Evaluation of an Onboard Real-Time Fault Management Support System for Next-Generation Space Vehicles

Intelligent Spacecraft Interface Systems (ISIS) Lab
Human Systems Integration Division
NASA Ames Research Center
Moffett Field, CA 94035

May 18, 2006

Contributors

Brent R. Beutter, Ph.D.
Michael Matessa, Ph.D.
Jeffrey W. McCandless, Ph.D.
Robert S. McCann, Ph.D.
Lilly Spirkovska, Ph.D.
NASA Ames Research Center
Moffett Field. CA

Dorion Liston, Ph.D.
Miwa Hayashi, Ph.D.
Valerie Huemer, M.S.
Joel Lachter, Ph.D.
Ujwala Ravinder, M.S.
San Jose State University
Moffett Field, CA

Steven Elkins, B.S.
Fritz Renema, B.S.
QSS, Inc.

Captain Robert Lawrence, United Airlines (Retired)
Battelle Research Institute
Mountain View, CA

Andrew Hamilton
NASA Johnson Space Center
Houston, TX

Acknowledgments

*This work was supported by the Engineering for Complex Systems (ECS) Resilient Systems and Operations Program (Subproject 4.1.5, Define User Interface Requirements for Advanced Health-Management Systems on Next-Generation Spacecraft)***,** *and by Grant # 01-OPBR-07-0000-0164 (Development of a Crew-Systems Concept for Vehicle Health Management in Next-Generation Spacecraft), from the NASA Space Human Factors Engineering project. Ames Research Center/ Dr. Steven Zornetzer provided additional funding. We also wish to acknowledge support from Human Factors Research and Technology Division Chiefs Terry Allard and Patty Jones, Dr. Mary Kaiser, and contracting managers Kevin Jordan and Marlene Hernan.*

Bruce Hilty of NASA Johnson Space Center (JSC) provided ISIS human factors researchers with the opportunity to learn about and make a career of spacecraft operations. He is just as responsible for the existence of the ISIS lab as anyone on the list of direct contributors. Other JSC personnel who contributed to the development of the lab and Fault Management Support System (FAMSS) design concepts include Tahqiq Abbasi, Don Berryman, Kelsey Berumen, Brian Bihari, Karsten Braaten, Travis Carter, Benson Chang, Francis Choi, Stacey Crear, Danny Deger, Steve Everett, Steve Gauvain, Terry Gobert, Ken Hill, Howard Hu, Gregory H. Johnson, Jennifer Madsen, David McDill, Mason Menninger, Bryan Miessler, Lee Morin, Victoria Palmer, Gene Peter, Michael Rosburg, Bohdan Scharunovych, Amy Shinpaugh, Ray Sims, Tom Smith, Kamila Smolij, Shashi Srinivasa, William Stahl, Kevin Taylor, Hadi Tjandrasa and Brian Ulczynski.

Thanks to Melissa Medina for timely, competent administrative assistance.

Thanks to Dr. Leland Stone, whose oculometrics lab provided crucial development and validation of many of the eye-movement processing and analyses tools used here.

Special acknowledgements to our 14 United Airline pilot participants (the "Magnificent 7" and the "Space CAUboys"), who have given so much of their time and attention to this project. We (reluctantly) withhold their names only to protect the integrity of their data.

Finally, the ISIS lab would not exist without the support of the late Dr. Joan Pallix, who read our early papers, believed in the concept of a human factors spacecraft simulator at NASA Ames Research Center, and created an ECS program element to develop and nurture it.

Table of Contents

Figures

Tables

Executive Summary

Human-rated spacecraft contain very complex and often highly interconnected engineering systems that must perform to precise operational specifications in very harsh environments. Critical systems are instrumented with sensors that provide real-time numeric readings of operating parameters. If a predetermined number of consecutive sensor readings fall outside the range consistent with normal (nominal) system operations, crew and ground personnel are alerted to the problem and the cause must be identified. If the cause is determined to be a genuine system malfunction (rather than, for example, sensor failure), the appropriate recovery procedures must be accessed and executed. Because malfunctions in the more dynamic systems can pose an immediate threat to crew safety or mission success, the crew must work the procedures to restore critical system function as quickly and accurately as possible.

Real-time fault management – the process of detecting, isolating and recovering from systems malfunctions – is one of the biggest operational challenges facing the crews of today's space shuttles. The shuttles' caution & warning (C&W) systems primarily use bounds checking methods to determine off-nominal performance; the crew is not aware of the potential existence of a malfunction until a threshold is reached. Moreover, off-nominal performance in one component often leads to a cascade of off-nominal performance in interconnected components, presenting the crew with a potentially large set of C&W events that they must associate with the signature of a single fault. This root-cause determination is further complicated by cockpit avionics and display limitations. Only a fraction of the sensed data is available on cockpit displays and an even smaller fraction can be viewed at once. To make matters worse, the display formats themselves are often poorly organized and highly cluttered, taking the form of closely-spaced matrices of digital data that may require considerable mental translation to understand the current operational status or functional mode of a system. Once the root cause is determined, operational challenges continue through the isolation and recovery activities. Malfunction-recovery procedures are only available in paper checklists. Beyond the purely psychomotor issues of accessing the checklists when crewmembers are fully suited and restrained in a vibrating vehicle, checklist navigation is inherently complex. The crewmember must locate the correct procedure for the root cause, navigate through the checklist steps by deciphering specialized symbols, abbreviations, boundary delimiters and spatial configurations, evaluate logical expressions by referring to other cockpit instruments and displays, perform mode reconfigurations by finding and toggling the correct switches from the hundreds of manual switches that populate the interior, and ensure that all steps are completed accurately and that the resulting system state is as expected.

Designers of next-generation crewed exploration space vehicles have three decades of technology advances at their disposal to reduce fault management difficulty and streamline fault management operations. Integrated System Health Management (ISHM) technologies can facilitate the process of detecting and isolating faults. Some of these technologies have already been incorporated in a prototype Enhanced C&W system for shuttle. Advanced navigation schemes for electronic checklists can facilitate the process of executing recovery procedures. Lastly, Human Factors and Human-Computer Interaction technologies can facilitate the process of organizing needed information and presenting it so that it better supports the crew's fault

management tasks. As part of a shuttle cockpit avionics upgrade (CAU) program, human factors researchers and shuttle crewmembers have already developed prototype displays incorporating some of these display improvement techniques.

In this report, we describe a concept that integrates these technologies into a FAult Management Support System (FAMSS) that assists the crew with all aspects of real-time fault management, from fault detection and crew alerting through fault isolation and recovery activities. The FAMSS concept specifies an intermediate level of crew-FAMSS functional allocation and user interfaces to enable and support that allocation. FAMSS automatically performs root-cause analyses, evaluates checklist logical expressions and makes switch throws. The crew maintains overall authority and control over the fault management process because FAMSS does not execute any procedure until a crewmember gives it permission to do so. This proposed functional allocation is enabled and supported by a FAMSS user interface, the Fault Management Display, which combines C&W and electronic checklist interface design features with CAU display format principles. The Fault Management Display is divided into two sections: a localized system schematic and an area for written (text-based) fault management procedures. Where possible, procedure information is coded into the system schematic, providing a graphics-based (as well as text-based) depiction of the procedure to assist the crewmember in understanding the required system reconfigurations and their effect on system function.

We recently completed an extensive empirical evaluation of FAMSS in the Intelligent Spacecraft Interface Systems (ISIS) laboratory at NASA Ames Research Center. Fourteen highly experienced commercial airline pilots assumed the role of spacecraft operator during the launch and ascent phase of eight spacecraft missions in a part-task (single-operator with no ground support) reconfigurable cockpit simulator. The baseline condition for the evaluation combines the shuttle C&W system, a CAU display suite, paper checklists, and manual switch throws. The FAMSS condition automates switch throws and root-cause determinations, removes the need to consult paper checklists (by providing checklist steps on the Fault Management Display), and adds the Fault Management Display to the CAU display suite. The evaluation methodology combined the standard suite of human performance measurement tools – accuracy and latency performance measurements, and situation awareness and workload questionnaires – with two infrequently used methods – eye movement analyses and predictive modeling of human performance.

The variety of evaluation techniques revealed many FAMSS-related empirical benefits to on-board fault management. Working malfunctions in conjunction with FAMSS assistance improved malfunction resolution accuracy by 43% and reduced malfunction resolution time by 54%. FAMSS reduced or eliminated fault management errors in a wide variety of fault management activities, including root-cause determinations of clusters of C&W events, reading and navigating through the checklists, and manually throwing switches. Similarly, determining root cause and navigating to the appropriate recovery procedure took much longer when FAMSS was not available.

FAMSS also greatly reduced participants' subjective perception of workload. Participants rated their workload on off-nominal (malfunction-containing) runs as 27% to 37% lower in the

FAMSS condition than in the baseline condition, with perceived greater benefits of FAMSS' automation in higher-complexity fault management situations.

If not carefully designed, automation can lead to significant decreases in situation awareness. The FAMSS concept specifies an intermediate level of automation to alleviate this potential problem, and the results indicate this goal was met. Objective situation awareness questions showed that participants' understanding of the environment was approximately identical for the Baseline condition compared to FAMSS. Subjective situation awareness results were stronger, with the ratings indicating that the participants actually increased their perceived ability to diagnose and resolve the malfunctions.

A secondary goal of the FAMSS evaluation was to determine the benefits and deficits of an integrated evaluation methodology that blended eye movement and predictive modeling methods with analyses of traditional human performance metrics such as response time and accuracy. Eye movement data augmented the data collected by traditional means in various ways. Eye movement tracking enabled us to gather independent evidence that helped clarify or deepen our understanding of how participants utilized critical features of the Fault Management Display. In particular, eye movements show that participants generally crosschecked schematic and text-based representations of procedures on the Fault Management Display. This suggests that participants found the embedded graphical depiction of procedure steps beneficial. Further, analyses of eye movements showed that participants return to their normal methodical instrument scan more quickly when FAMSS provided fault management assistance. This corroborates the improvements suggested by subjective situation awareness ratings.

In addition to assessing the many benefits of the FAMSS concept, the evaluation revealed two potential drawbacks with the FAMSS interface. First, FAMSS provided little information on failure impacts. More explicit information would alleviate problems of mistaking a propagated "daughter" fault as a bona fide fault. Second, the interaction between crewmember and FAMSS could be clarified in the case of multiple malfunctions. Some of the participants expected FAMSS to automatically switch to the next malfunction to work when the current malfunction procedure was completed. This is not a feature of the current concept. Pending tasks need to be more clearly depicted and perhaps reminders provided.

The shuttle operations paradigm has been refined over 25 years of flight. Each task that the crew is required to accomplish onboard is developed, perfected, and practiced many times before flight. Simultaneously, ground controllers also learn, practice and perfect their tasks of systems monitoring and failure diagnosis. Though it may be desirable to reduce training time or introduce automation to lessen crew and ground controller workload, for the most part, the paradigm works well and leads to successful missions. Nevertheless, the circumstances of next-generation vehicle missions will require the crew to operate their vehicles in a more autonomous (independent) mode than they do today. A fault management support system could provide invaluable assistance under these conditions.

1 Introduction

1.1 Background: Systems Malfunctions and Manned Spaceflight

Human-rated spacecraft contain very complex and often highly interconnected engineering systems, including propulsion systems; electrical and mechanical power generation and distribution systems; guidance, navigation, and control (GN&C) systems; data processing systems; life support systems; and communications systems. Particularly during the dynamic phases of a mission, such as launch, ascent, and entry, these systems must perform to precise operational specifications in very harsh environments, whose cumulative effects on system functioning are often poorly understood. As a result, systems malfunctions are an ever-present threat to mission success and crew safety.

The risks posed by systems malfunctions influence almost all aspects of a manned spaceflight program, from the initial stages of vehicle design all the way through real-time mission operations and vehicle maintenance. For their part, designers of spacecraft systems reduce malfunction risk by building in functional redundancies that provide opportunities to diagnose and understand malfunctions and restore safe systems operation. Systems engineers carry out failure modes and analyze operational data to identify a wide range of possible systems malfunctions, understand their impact on system performance and functionality, and determine how to manage systems redundancies to minimize the impact of the malfunction (safe the system) and, where possible, restore critical functionality. This knowledge is then captured in the form of malfunction-specific procedures that specify the sequence of activities that should be taken by the crew in the event a malfunction occurs during flight.

During an actual mission, monitoring, managing, and maintaining the health of vehicle systems accounts for a significant fraction of real-time mission operations. Each system is instrumented with sensors that provide real-time numeric readings of critical operating parameters, such as temperatures, pressures, accelerations, and flow rates. If a predetermined number of consecutive sensor readings fall outside the range consistent with normal (nominal) system operations, crew and ground personnel must be alerted to the problem and the cause must be identified. If the cause is determined to be a genuine system malfunction, the appropriate procedures must be accessed and executed. Particularly during the dynamic phases of flight, malfunctions in the more dynamic systems can pose an immediate threat to crew safety and mission success. Thus, there is a strong need to work the procedures to restore critical system function as quickly and accurately as possible.

Unfortunately, a shuttle crew's ability to perform these fault management activities from the cockpit (i.e., without ground assistance) is compromised by several factors. Many of these factors stem from the limited capabilities of the onboard data processing and vehicle health management technologies, some of which date from the 1970s. The next section provides a brief overview of these factors and how they impact crew performance.

1.2 Fault Management in Today's Spacecraft: Human Factors Issues

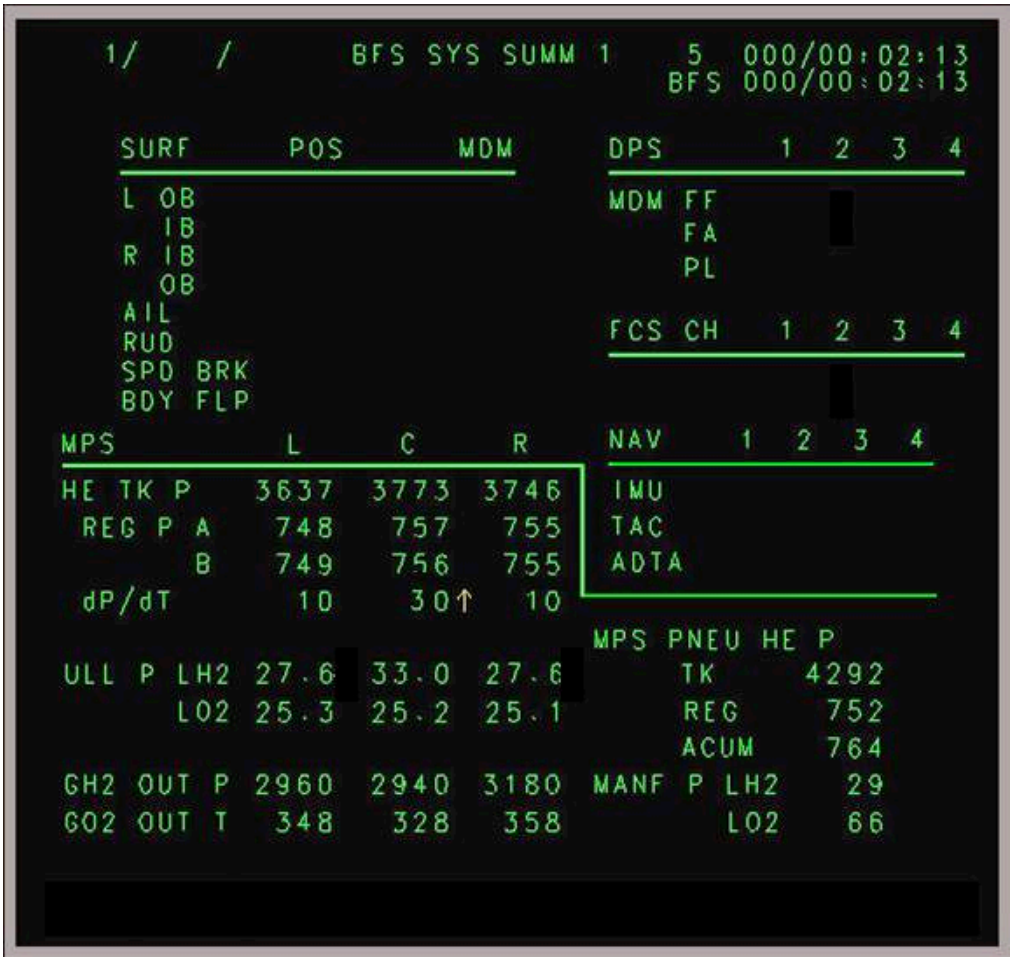


Figure 1.1. *BFS GNC SYS SUM 1 Display (MEDS Cockpit). Note the "up" arrow beside the off nominal dP/dT reading for the C (Center Engine) helium supply system.*

Real-time fault management can be a very complex task, encompassing a large number of information processing components and physical activities that require a variety of operational skills and forms of knowledge. Organizing, scheduling, and coordinating these components place high demands on crewmembers' attention and working memory resources. If these resources are overloaded, or if there is a skill or knowledge deficiency in any one of the task elements, operational errors can result with potentially catastrophic results. In this section, we describe some of the factors that contribute to the complexity and difficulty of onboard fault management.

1.2.1 Availability and Display of Systems Information in the Cockpit

Safe and effective fault management is critically dependent on crewmembers having a good understanding of the current operating mode of the malfunctioning system, how fault isolation and recovery procedures will alter that mode, and what impact the altered mode will have on system functionality and capability. Ideally, all the information necessary to support this understanding would be accessible on cockpit displays in a form that is quick and easy to assimilate. In the shuttle cockpit, both the display and processing of systems information is

compromised by limitations of the onboard avionics and data processing systems and other factors. Only a fraction of the sensed data is even available on cockpit displays. Of that fraction, display real estate limitations dictate that an even smaller fraction can be viewed at any one time. To view all available data about a system, a crewmember often must navigate through several successive display formats. Interpreting the display formats is sometimes quite challenging because many formats are quite cluttered, taking the form of closely-spaced matrices of digital data that may require considerable mental translation to understand the current operational status or functional mode of the system. An example of a densely populated numeric display is the Backup Flight System (BFS) Guidance, Navigation, and Control (GNC) System (SYS) Summary (SUM) 1 display, shown in Figure 1.1. BFS GNC SYS SUM 1 blends guidance, navigation and control information towards the top section with critical main propulsion system operating parameters towards the bottom. The green line snaking through the middle of the display roughly demarcates the two sections.

1.2.2 Caution and Warning System

Like many flight vehicles, the shuttles are equipped with a Caution and Warning (C&W) system whose function is to alert the crew to the presence of an anomalous operating condition and to provide relevant information to help them diagnose the source of the problem. The system consists of electronics (hardware) and software that provide the crew with visual and aural cues when automatically monitored parameters exceed preset limit values. The crew interfaces with the primary (hardware) C&W system through a message annunciator matrix, a parameter status light matrix, four (interconnected) red MASTER ALARM pushbutton indicators and aural alarms. The crew interfaces with the backup (software) C&W system through one element of the (hardware) message annunciator matrix, the MASTER ALARM pushbutton indicators, aural alarms and fault messages displayed flashing on the cockpit displays. The text of the fault message identifies the system where limits are being exceeded. It is also frequently used as the title of the flight data file procedure to be worked by the crew, as discussed in the next section. Because the software processes each sensor's data independently, the C&W system cannot discriminate a legitimate off-nominal reading from a spurious reading due to either a failed sensor or a failure in a digital signal conditioner or other component of the shuttle's data processing system. More seriously, due to the complex and often highly interconnected nature of the onboard systems, a failure of one component frequently causes additional abnormal sensor readings and changes in the operational status of subsystems and equipment downstream of the instigating failure. When these forms of failure propagation occur, the result is often a cascade of C&W events (where each event is composed of a combination of an auditory and visual alarm and accompanying fault message) that distract the crew, impair their situation awareness, and hamper their ability to determine the root cause of the problem (McCandless, McCann, & Hilty, 2003).

1.2.3 Paper Flight Data Files

Within the shuttle cockpit itself, the procedures for isolating and recovering from systems malfunctions are available only on cue cards or in paper documents called flight data files. From a purely psychomotor perspective, accessing the information in a flight data file (FDF) can be difficult, particularly during dynamic flight phases when crewmembers are suited, restrained, and wearing helmets that restrict their effective field of view, and the cockpit is vibrating. Each FDF is organized into sections, one for each significant system on the vehicle. A paper tab containing a system identifier or related information (e.g., MPS for Main Propulsion System) is appended to

section pages to help the crewmember locate specific systems or malfunctions. Each section contains several malfunction-specific checklists of instructions and procedures. Each checklist begins with a main procedure title, written in boldface. In many cases, the title corresponds directly to a software C&W fault message. Thus, the crew first processes the C&W events. If there are multiple fault messages, they select the message that corresponds most closely to a root cause, and then locate its match in the appropriate section of the appropriate FDF.

The main procedure title heads a subsection containing the off-nominal instructions that apply to a particular system or subsystem. Once the subsection is located, the crewmember starts navigating through the instructions, which are coded in the form of specialized symbols, abbreviations, boundary delimiters, and spatial configurations that collectively require extensive training to decipher and understand (Figure 1.4 shows an example set of procedures). Individual instructions frequently take the form of logical expressions (IF-THEN-ELSE statements) that crewmembers must evaluate by locating and processing systems or flight status information on cockpit instruments and displays. The outcome of the evaluation of the logical expression determines which path should be taken through the rest of the section. That, in turn, determines what instructions have to be carried out, and in what order.

Checklist navigation is inherently complex. Many C&W fault messages (and hence, many FDF procedure titles) are to some extent generic (often because the underlying instrumentation is sparse or nonexistent). For example, they may point to a likely leak in a system, but not the precise location of the leak (i.e., not the precise tank, feedline, or manifold that is experiencing the leak). In these cases, the instructions may first designate a procedure to reconfigure the operating mode and examine the way the system responds to the reconfiguration. Depending on the particulars of that response, the location of the leak may be revealed. At that point, the appropriate remedial action can be applied (e.g., isolating the affected component from the rest of the system). If not, additional actions may be required to determine the location of the problem and select the appropriate procedures for dealing with problems in another location. An example of this kind of redundancy management and associated FDF navigation requirements is described in Section 1.4.2.

1.2.4 Manual Mode Reconfigurations

First-time viewers of the shuttle cockpit frequently comment on the hundreds of manual switches that populate the interior. These switches control the operational mode of the onboard systems; for example, valve-control switches typically have three positions: Open, Closed, and GPC (which places the valve under General Purpose Computer [GPC] control). Switches are typically organized by switch control panel and then by system. For example, most main propulsion system switches are located on switch control panels to the right of the pilot's (right-seat crewmember) seat; many GPC switches are located on an overhead control panel more accessible to the commander (left-seat crewmember) than the pilot; and many environmental control and life support system switches are on switch panels to the left of the commander's seat. The high density of switches on these panels is partly the result of design redundancies. For example, consider the subsection of switch panel R2 shown in Figure 1.2. This section, which is only the top left corner of the actual panel, contains the switches for the Main Engine helium supply systems. The top row contains the LEFT, CTR, and RIGHT engine helium isolation leg A ("He ISOLATION A") valve switches; the second row contains the corresponding helium isolation leg B ("He ISOLATION B") valve switches; and the bottom row contains the LEFT,

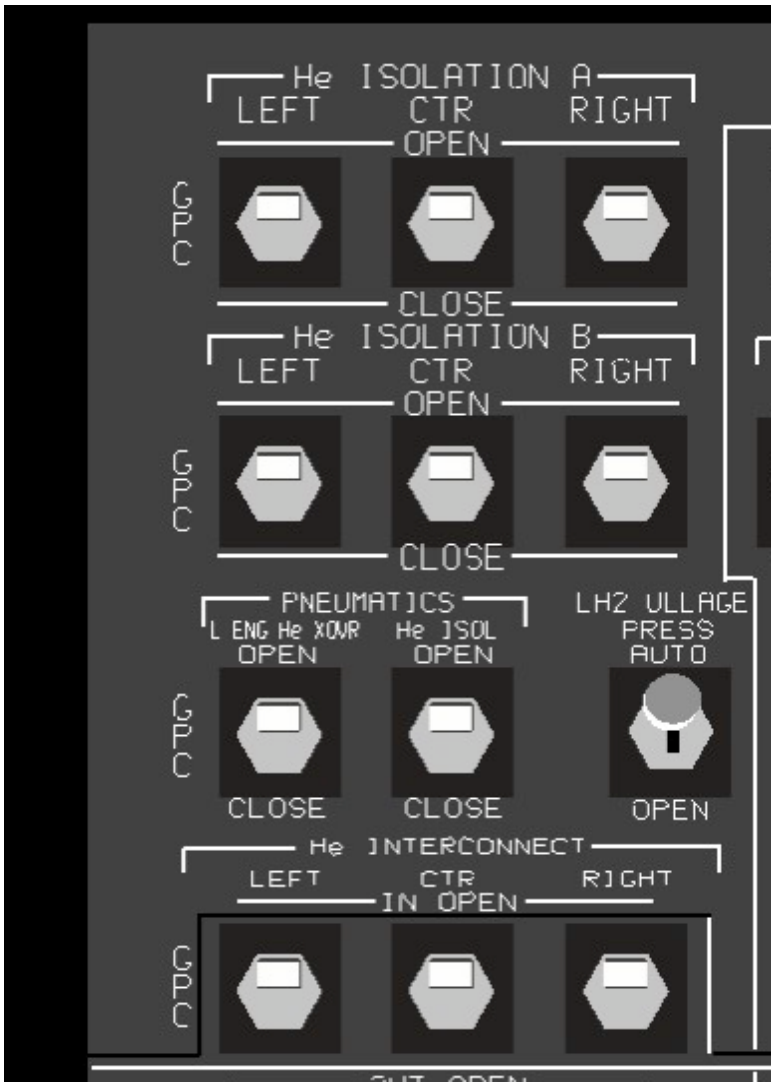


Figure 1.2*. Upper left Corner of Switch Panel R1, showing the switches for the three main engine helium supply systems. Left Engine switches are in the left column; Center Engine switches in the center column; Right Engine switches in the right column.*

CTR, and RIGHT engine common manifold (He INTERCONNECT) switches. We will have much more to say about the helium supply systems and the functions of these switches in Section 1.4.2. For now, the point we want to make is that every time an instruction in the FDF calls for a switch throw, the crewmember must remember the location of the appropriate switch panel, locate and attend to the appropriate switch in a dense field of visual similar stimuli, and manually toggle the switch to the commanded position.

As with the paper FDF's, physical access to cockpit switches can be difficult during the dynamic flight phases, when crewmembers' mobility and reach are restricted by spacesuits, gloves and helmets. In addition to these purely psychomotor problems, the densely cluttered environment of virtually identical switches can make it difficult to attend to, and select, the correct switch. As with any physical action that involves selecting a target element from perceptually similar distractors, there is always the possibility of a "slip" (Reason, 1990) – an error in executing a motor command that leads to an unintended action, such as inadvertently toggling a switch adjacent to the intended target. Assuming the correct switch has been selected and toggled, the crewmember must ensure that he or she has moved the switch to the commanded position and accomplished the desired mode transition. In many cases, hardware "talkback" indicators, adjacent to the switches, provide visual feedback as to the updated configuration, e.g., whether the valve controlled by the switch is open, closed, or in transition. In other cases (including the helium supply switches in Figure 1.2), there are no talkback indicators, so the results of a switch throw have to be inferred from changes to numeric parameters on displays such as BFS GNC SYS SUM 1. Often, these indicators are quite subtle.

1.2.5 Human Factors Risks with Real-time Fault Management: Summary of Impacts

We noted earlier that design redundancies play a central role in reducing the risk posed by systems malfunctions. However, from an operational perspective, redundancy provides no risk reduction whatsoever unless it is managed effectively. On current generation spacecraft, redundancy management entails a variety of information processing requirements (e.g., FDF navigation) and manual mode reconfigurations (e.g., switch throws) that are highly labor-intensive and provide multiple opportunities for human error.

Even when fault management operations are performed correctly, the various difficulties and human factors problems inflate the *time* required to work through and complete fault management-related operations, to the point where some malfunctions take minutes to resolve. Since focal attention is required for most of these activities (such as reading and navigating the FDF and finding, throwing, and confirming cockpit switches), a crewmember's nominal scan of instruments and displays is either greatly disrupted or eliminated. A recent analysis of astronaut scanning patterns on nominal ascents revealed that in a single-operator environment, these highly trained operators devote most of their attention to mission or flight-related information sources, such as their primary flight instruments (Hayashi, Huemer, & McCann, 2005; Huemer, Matessa, & McCann, 2005). The longer a crewmember's attention is diverted from these nominal sources, the greater the potential loss of vehicle and mission-level situation awareness.

In addition to the danger posed by this cognitive tunneling, it is a simple fact that the longer it takes crewmembers to resolve a malfunction, the greater the risk that another malfunction will occur before the current problem has been resolved. In that event, crewmembers are suddenly forced with timesharing fault management activities. The high workload and attentional requirements of the constituent operations leaves little spare capacity to work additional malfunctions. Indeed, given the focused demands of most constituent activities, a single crewmember would have to perform many of them in a strictly serial manner, delaying remedial operations and leaving systems in off-nominal operating modes with mission and safety impacts that may be growing over time.

But overlapping malfunctions have many opportunities to decrease crew response to a malfunction beyond delays caused by serial operations. Annunciation of the second problem interrupts the handling of the first, and crewmembers must, at a minimum, determine whether the new problem has a higher priority than the existing one by switching their attention to the newly-annunciated C&W event(s), extinguishing the alarms and reading the fault messages. Prioritization could be complicated and time consuming, delaying resumption of the interrupted activity. But humans are notoriously inefficient when it comes to resuming interrupted tasks; they forget where they were in a checklist and possibly repeat or omit steps, etc., further delaying resumption of the interrupted activity. For a variety of reasons, then, the temporal impact of slow and inefficient fault management operations during dynamic flight pose risks to mission safety over and above the risks associated with performing the mental and physical operations required to handle each malfunction individually. The temporal impact of switching contexts between dealing with an existing problem and an overlapping new problem is just one very specific example of why malfunction handling system designers should attempt to minimize the chances of multiple-malfunction tasking situations as much as possible.

1.3 Fault Management on Next-Generation Space Vehicles: Opportunities

Because the space shuttle was designed in the early to mid-1970s, designers of the next-generation of crewed exploration space vehicles have three decades of advances in portable computing power, information processing technologies, and human-centered interface design at their disposal to reduce fault management difficulty and streamline fault management operations. In this section, we review these advances, beginning with those that have occurred within the shuttle program itself.

1.3.1 Cockpit Avionics Upgrade (CAU) Display Formats

In the 1990's, NASA initiated a hardware upgrade of the original 1970's-era orbiter cockpits to replace the original electronic cathode ray tubes ("green screens") and mechanical flight instruments with liquid crystal displays (LCDs) driven by dedicated processing devices. This upgrade, called the Multifunction Electronic Display System (MEDS) cockpit, was first flown in Space Shuttle Atlantis in 2000. NASA then approved a software-oriented Cockpit Avionics Upgrade (CAU) project, whose charter was to exploit the enhanced display and computing capabilities of the MEDS cockpits to address human factors problems with the legacy display formats (most of which were "ported" directly to glass) and other operational problems in the cockpit. The new display formats consolidated information from over 100 "green screen" formats, such as BFS GNC SYS SUM 1, onto fewer display formats, greatly reducing the need for display navigation. Moreover, many of the redesigned formats incorporate schematics and other graphical features not present in the originals, particularly on the redesigned system summary displays that provide information about system health and operating mode. Since a major focus of this the report is on how participants manage faults in the helium supply systems for the shuttle's main engines, we illustrate these graphical features with the CAU Main Propulsion System Summary (MPS SUM) Display. Before describing those features, a short segue into the function and architecture of the helium supply systems is necessary.

Function and Architecture of the Main Engine Helium Supply Systems. During the powered flight phase of a shuttle mission, which lasts from liftoff to approximately 8.5 min of Mission Elapsed Time (MET), each of the three main engines is supplied continuously with gaseous helium. The helium is used to pressurize an intermediate seal in the engines' high-pressure oxidizer turbopumps, which boost the pressure of the liquid oxygen (LOX) before it is injected into the main combustion chamber. The purpose of the seals is to prevent LOX from mixing with the fuel-rich hot gas that drives that turbopump. Mixing of the propellants in this region of the main propulsion system could lead to a catastrophic explosion. In fact, the danger posed by the mixing of these volatiles is so great that, if the pressure in the seal falls below a critical threshold, the engine shuts down automatically, even though the shutdown brings on a mission abort situation that is quite risky in its own right.

Each engine has a helium supply stored at high pressure in a dedicated set of storage tanks. The simplest design would have been to connect the tanks to the turbopump through a single feedline. However, a leak in the feedline could deplete the helium supply before the scheduled main engine cut-off time, leading to premature engine shutdown and mission abort. Main Propulsion System (MPS) designers addressed this "single point" failure vulnerability by splitting the helium flow into two redundant feedlines, Leg A and Leg B, and outfitting each leg with its own pressure regulator and isolation valve. During nominal system operations, both isolation valves

are open, so helium flows through both legs simultaneously. However, if a leak develops in one leg or the other (or if the Leg A or Leg B pressure regulator fails), the crew has the option of simply closing the affected leg's isolation valve. This simple reconfiguration either isolates the leak, or prevents helium from flowing through the failed regulator, while maintaining nominal flow through the other leg.

Depictions of helium supply systems on MPS SUM. Shown in Figure 1.3, MPS SUM graphically depicts the helium supply systems (backup, left, center, and right engine, respectively) along the top tier of the display. The rectangular boxes represent the helium storage tanks for each system. The connectors depict the supply lines from the tanks to the engines, including the split into legs A and B. Both legs are broken by circles, also labeled A and B. These circles, together with the segment of the supply line that is embedded inside, symbolize the helium isolation valves. When the valves are open (nominal configuration), the segment inside the circle is aligned with the rest of the supply line, and both line and circle are rendered in bright white, signaling flow. When a valve is closed (as is ISOL A on the Right Engine helium supply system in the figure), the interior segment of the valve symbol is rotated 90 degrees with respect to the rest of the feedline, breaking perceptual continuity, and both the valve and the feedline segment underneath are rendered in dark gray (signaling no flow). When a valve is failed closed, as is ISOL A in the Center Engine supply system, the valve symbol is red.

*Empirical Evaluation*s *of the CAU Display Formats*. These graphical features and system configuration codes provide a more intuitive depiction of the helium supply systems, their current configuration mode, and their operational status, than the rows and columns of numeric values on BFS GNC SYS SUM 1 (Figure 1.1). Similar graphical depictions of systems architecture and operating mode were incorporated into the new system summary displays for many other orbiter systems. Recently, a thorough astronaut-in-the-loop empirical evaluation of the redesigned formats (along with other CAU upgrades, such as some entirely new display formats) on fault management capabilities during dynamic flight phases was carried out in a high-fidelity (full-mission) simulator at NASA Johnson Space Center (JSC). The CAU formats improved crewmembers' fault management resolution time and accuracy, reduced their workload, and enhanced their situation awareness (as measured by both objective questions and subjective ratings), compared to when they worked the identical malfunctions in the MEDS cockpit (McCandless, McCann, Berumen, Gauvain, Palmer, Stahl, et al., 2005; see also Hayashi, et al., 2005, for similar results in part-task simulation).

Partially due to budget constraints, the CAU project was cancelled in 2004 before the upgraded display formats could be implemented in the shuttles. However, prior to the cancellation, NASA embraced a Vision for Space Exploration (VSE) that calls for the development of a new generation of crewed spacecraft to support missions beyond low-earth orbit. The agency's commitment to build a new generation of space vehicles places the CAU display redesign effort in an entirely new light. CAU displays represent the state of the art in spacecraft cockpit display format design, and their value has been thoroughly established in ground testing; as such, they are the logical departure point for the design of cockpit display formats on VSE vehicles. That assumption is central to the work that follows.

1.3.2 Integrated System Health Management Technologies: Enhanced C&W

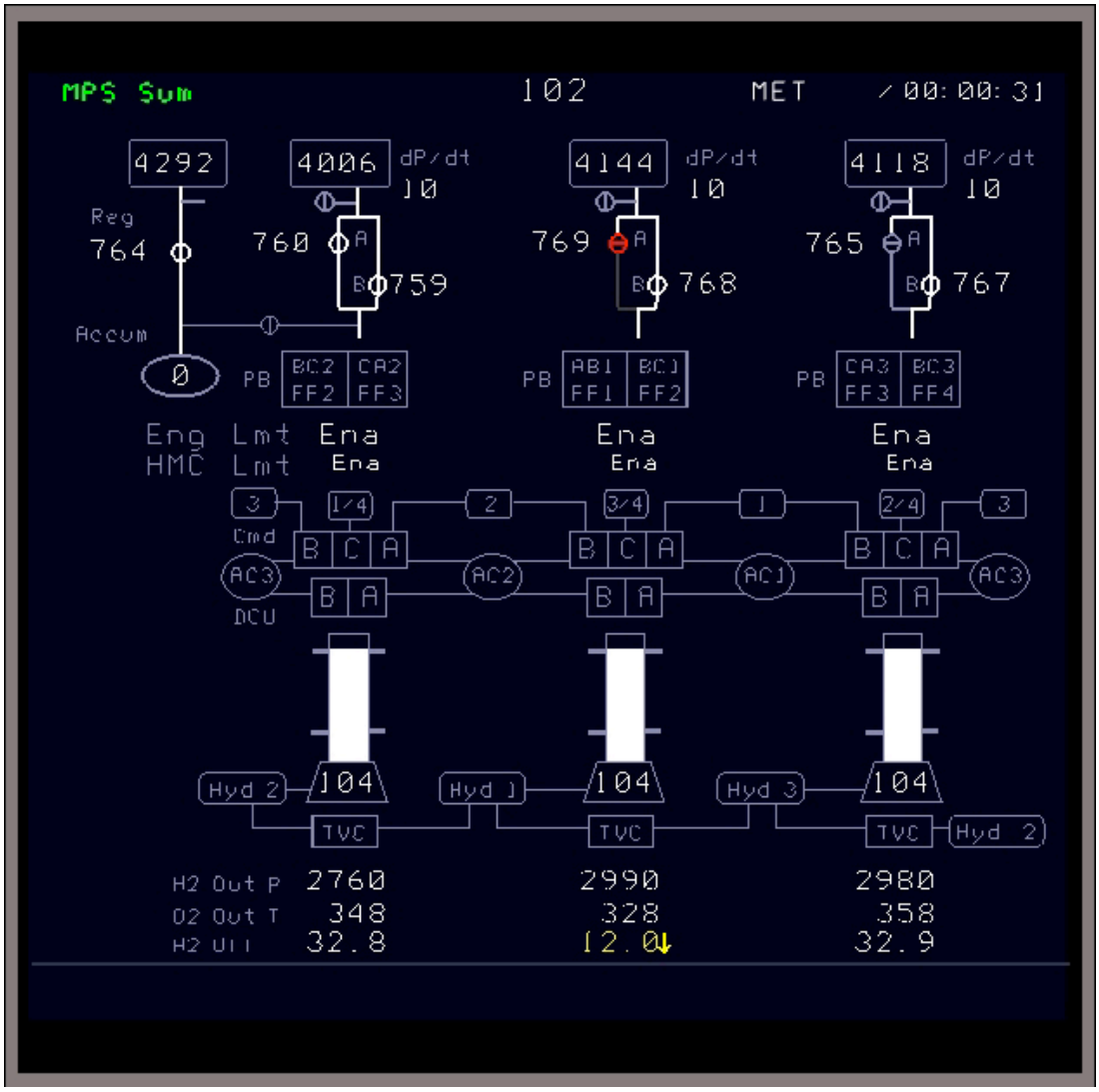


Figure 1.3. *Cockpit Avionics Upgrade Main Propulsion System Summary (MPS SYS SUM) Display. The top row depicts the Left, Center, and Right Engine Helium Supply Systems, and the Backup (Pneumatic) System on the far left. Helium tank pressures are indicated by the digital values inside the rectangular boxes; dP/dT by the digitals beside the tanks; and Leg A and Leg B regulator pressures by the digitals beside the isolation valve symbols. The figure shows Center Engine ISOL A failed closed, Right Engine ISOL A nominally closed, and the remaining isolation valves nominally open. The bottom row of the display contains ullage pressure readings for the liquid hydrogen (aft) external tank and an off nominal low reading from the Center Engine ullage pressure sensor.*

Along with identifying human factors problems with the existing cockpit display formats, the CAU project identified the shuttle C&W system as another important source of human factors and usability problems, and assigned a team to address those issues. Consequently, in parallel with the cockpit display redesign efforts, CAU project members developed a detailed concept, and delivered much of the supporting software, for an Enhanced Caution and Warning (ECW) system (for details, see McCandless et al., 2003). Essentially, the ECW project focused on building a sophisticated filter for C&W events that takes as input temporal clusters of C&W events, applies failure identification logic to make "root cause" failure determinations for the

clusters, and allows only the C&W event associated with the root cause to annunciate in the cockpit (i.e., all events associated with downstream consequences of the instigating failure are inhibited). Additionally, root cause events that are not immediately significant during a dynamic flight phase are suppressed until a lower workload flight phase.

ECW was not incorporated in the human-in-the loop evaluations of fault management performance in the CAU cockpit (McCandless, et al., 2005; Hayashi, et al., 2005), so the precise impact of C&W event filtering is not known. Even without a formal evaluation, however, there seems little doubt that ECW would improve crewmembers' ability to quickly focus attention on an off-nominal condition and determine the root cause of the problem.

Beyond progress in the shuttle program itself, in the decades since the shuttles were first designed, the field of applied artificial intelligence known as Integrated System Health Management (ISHM) has advanced to the point where machine-based (automated) systems managers have been built with "end-to-end" fault detection, isolation, and recovery capabilities. These systems generally employ sophisticated pattern recognition technologies (e.g., statistical analysis, neural networks, fuzzy logic, data mining) that continuously classify real-time sensor readings as being consistent with either nominal or off-nominal modes of systems functioning. When an off-nominal pattern is detected, a "reasoner" (e.g., rule-based expert system, case-based reasoning system, model-based reasoning system, learning system, or probabilistic reasoning system) then determines the root-cause based on that pattern. The failure diagnosis is then passed to a "reactive planner" that 1) determines the procedures required in order to achieve the desired goal state (typically, a return to nominal functionality), 2) determines the correct sequence of systems reconfigurations (procedures) needed to achieve that goal state, 3) physically commands the procedures, and 4) processes sensor data to determine whether the action(s) have been carried out and the desired mode reconfiguration(s) has been achieved.

In the years prior to the announcement of the VSE, NASA supported several high profile programs to replace or augment the shuttles with next-generation launch and transport vehicles. Many of these programs, most notably the Space Launch Initiative, yielded detailed concepts from aerospace contractors for highly integrated avionics system architectures, including data acquisition and data handling infrastructures to support real-time ISHM technologies. Meanwhile, NASA itself developed a detailed operations concept for next-generation spacecraft complete with detailed plans for how to combine these data processing capabilities with ISHM technologies to achieve more real-time fault management support than exists even on today's aircraft.

Advances from the shuttle's ECW project and more sophisticated ISHM technologies could both be used to help the crew determine the root cause of anomalous behavior. Which approach is used could depend on the complexity of the behavior. Further, ISHM technologies could be used to help generate the event inhibit and suppress rules necessary for an ECW-type approach for next-generation vehicles.

As we have seen, however, working a systems malfunction extends well beyond determining the root cause. In the very early stages of the CAU project, when human factors problems with the shuttle cockpit were formally identified and prioritized, the problem definition document for the

C&W system also called out the serious human factors issues with fault isolation and recovery activities, most stemming from the paper FDF's. The recommended solution was to convert the paper FDF's into electronic checklists, which could have been displayed in the MEDS cockpit.

1.3.3 Electronic FDF

Looking beyond the confines of NASA's human spaceflight programs, certain segments of the aeronautics industry have extensive experience in developing and implementing electronic versions of fault management checklists. The development phase stretches back to the early 1990s, when Boeing and Airbus began designing a new generation of state-of-the art glass cockpit passenger aircraft, which eventually became the B777 and A300 aircraft in widespread service today. Because these aircraft were built from scratch, designers were able to incorporate a much more integrated avionics system than on earlier aircraft. The new systems have unprecedented capabilities to share real-time data between historically stand-alone aircraft systems and data sources. One of the most notable changes enabled by this data sharing was the replacement of the traditional paper checklists of both nominal and off-nominal (emergency) procedures (the equivalents of the shuttle FDF's) with electronic versions. These versions are considerably more user friendly than the paper versions they replaced. With direct access to data on systems status and current operating mode, the onboard computers are able to evaluate many logical expressions in off-nominal checklists, shifting much of the computational burden of checklist navigation from crew to machine. Moreover, in part because the automated evaluation of logical expressions greatly reduces the number of possible navigation paths, electronic checklist designers were able to incorporate simple navigation cues that straightforwardly guide the crew through the correct sequence of procedures.

In addition, the highly integrated avionics allowed direct links between the crew alerting (caution and warning) system and the emergency checklist (ECL) system. Similar to the shuttle, the C&W fault messages were designed to be isomorphic with the main procedure titles in the ECL. This isomorphism enabled designers to use the fault messages as unique identifier codes in the emergency checklists databases, thereby automating the process of accessing and displaying the appropriate checklist to the crew.

Developers of next generation spacecraft are in a similar position to the commercial aircraft designers of the 1990's. Because these exploration vehicles are being designed and built from the ground up, there is a clear opportunity to incorporate a more integrated avionics system than on the shuttle. Just as on the B777 and A300 aircraft, such a system could enable the kind of comprehensive data processing and data sharing that supports dynamic versions of electronic flight data files, with all the display navigation simplification and streamlining these displays make possible.

1.4 Fault Management Support System (FAMSS)

Each of the information technologies and user interfaces identified in Section 1.3 could support a much more capable fault management system than the system available on the shuttle. However, we strongly believe that the utility of these technologies would be greatly enhanced if they were integrated into a single fault management system (Scandura and Garcia-Galan, 2004). Theoretically, such a system could incorporate CAU display design concepts, automated root-cause fault determination (a combination of ECW and ISHM), automated flight data file navigation (electronic checklists), and automated procedure execution (switch throws). In this

section, we describe a concept for an onboard Fault Management Support System (FAMSS) for VSE vehicles that integrates these technologies. FAMSS goes well beyond the functionality of the current C&W system, assisting the crew with all aspects of real-time fault management, from detection and alerting through isolation and recovery activities.

1.4.1 The FAMSS Concept

FAMSS is built on the assumption that VSE spacecraft will have four core capabilities. First, their glass cockpit displays will have color-coding and graphics capabilities necessary to generate CAU-style display formats. Second, automated root-cause determination will be provided by ECW-style failure identification logic, model-based reasoning, or some combination of the two. Third, onboard computers will have real-time access to all sensor data required to evaluate logical FDF expressions, automating the process of navigating through flight data file checklists. Fourth, off-nominal mode reconfigurations (switch-throws) are nominally performed by machine.

Following well-established principles for human-automation teaming (Malin, Schreckenghost, Woods, Potter, Johannesen, Holloway, & Forbus, 1991), a highly automated concept such as FAMSS must be designed to work in close coordination with the crew. Given the extensive functionality associated with FAMSS, we were concerned with the possibility of automating the fault management process to the point where crewmembers experience "out-of-the-loop" unfamiliarity problems (Endsley & Kiris, 1995) or lack understanding of automated activities and motivations due to clumsy human-automation interfaces (Billings, 1997). Hence, our approach to FAMSS design represents a compromise between two opposing considerations. On the one hand, we would like to automate as many of the activities as possible that contribute to the high workload, difficulty, and inefficiency of fault management operations today. The obvious but highly desirable goals are faster and easier operations with reduced opportunities for crew error. On the other hand, we were mindful of the painful lessons learned from 20 years of experience with over-automation and clumsy automation on commercial aircraft. In an effort to balance these considerations, FAMSS was designed to function at the intermediate level of four on a modified version of the well-known Sheridan-Verplank scale of human-machine function allocation (McCann and McCandless, 2003). When a vehicle malfunction occurs, FAMSS automatically performs a root-cause analysis and evaluates logical expressions in the flight data file, fully automating the process of flight data file navigation. At the other end of the fault management process, FAMSS takes care of system reconfiguration activities by automating switch throws. However, the crew still maintains overall authority and control because FAMSS does not execute any procedure until a crewmember gives it permission to do so. Additional details on the rationale behind our choice of crew/machine functional allocation can be found in McCann and McCandless (2003) and McCann and Spirkovska (2005).

Having established a candidate crew/FAMSS division of labor, the next step was to design a user interface to support it. Functionally, the "top-level" requirement for this interface was that it integrate traditional C&W functions of malfunction alerting and identification with Sheridan-Verplank Level 4 support for checklist navigation and procedure execution. From a human factors perspective, the "top-level" requirement was for a display that would support and enable good situation awareness of FAMSS intentions and FAMSS actions, and also of the changes to system mode and system functioning that would (and then did) come about as a result of those actions.

The standard approach to the design of electronic flight data files is to present procedures in the form of text-based instructions, essentially the same format as on paper. Following this convention, the FAMSS interface also presents FDF procedures in text form. In addition, however, we integrated design concepts developed by Malin, Kowing, Schreckenghost, Bonasso, Nieten, Graham, Fleming, MacMahon, & Thronesbery (2000) with CAU-based system schematics to, where possible, embed procedural information directly into a schematic representation of the system experiencing the malfunction. The idea was that graphical or schematic representations of complex physical systems more closely match experts' mental representation of these systems than text. When procedures are presented only in text format, some degree of mental translation is required in order to assimilate that information into the operator's mental model of current system configuration and what the configuration will be following the commanded mode transition (procedure). By redundantly coding procedural information within a system schematic, we hoped to assist with this translation, help verify the operator's understanding of the text command, or both.

We explain this "embedded procedures" concept more fully in Section 1.4.3. Before doing so, however, we want to first describe and explain the section of the Ascent-Entry Systems Procedures (AESP) document (the FDF containing the procedures for managing systems malfunctions during powered flight) that covers malfunctions in the main engine helium supply systems. Both the design and evaluation of the FAMSS concept were heavily influenced by the particulars of how participants manage malfunctions (such as leaks) in these systems, whose function was described earlier. The description will help illustrate the central role of redundancy management activities in many fault management operations, and to further illustrate the difficulties, complexities, and risks associated with paper checklist navigation. A more thorough understanding of these issues will help to better understand the changes that FAMSS brings to these operations and information processing requirements, and potential FAMSS impacts on fault management performance.

1.4.2 FDF Navigation for Helium Supply Systems Malfunctions

The procedures for working helium supply system malfunctions are found in the Main Propulsion System (MPS) section of the AESP FDF. Shown in Figure 1.4, the main procedure title, "MPS He P" (short for Main Propulsion System Helium Pressure) is also the fault message generated by the C&W system if, for any engine, the reduction in helium tank pressure in any three second period exceeds the amount consistent with the nominal flow of helium from the tank. Right away, therefore, the crew has to crosscheck the FDF instruction with BFS GNC SYS SUM 1 (Figure 1.1) to identify which helium supply system is experiencing the problem. The critical system parameter, labeled dP/dT (for change in tank pressure over time) is depicted on the line in the middle left-hand region of the display. The crew establishes which engine is experiencing the problem by checking these values directly (a check made easier by virtue of the fact that the C&W system inserts an "up"-pointing arrow beside the out-of-limits parameter). In Figure 1.1, it is the Center Engine supply system that is experiencing the problem.

If after MECO-60: The first logical expression. The next line in the checklist contains the first of several logical expressions that must be evaluated by the crewmember in order to navigate correctly through the remainder of the checklist. In this case, for example, if the off-nominal

MPS He P (Pre MECO)
1. Check dP/dT
If after MECO -60: 2. Shut dn MN ENG per MPS CMD/HYD/ELEC >>
If He REG P ↑ or ↓ : 3. (Aff) He ISOL – CL Otherwise: 4. (Aff) He ISOL A – CL If no decr in dP/dT: 5. (Aff) He ISOL A – OP B – CL If no decr in dP/dT:. 6. (Aff) He ISOL B – OP
If any ENG failed: 7. (Failed) ENG He I'CNCT – OUT OP
If nonisolatable: 8. Shut dn MN ENG per MPS CMD/HYD/ELEC If/when TK P < 1150 or REG P < 679: 9. (Aff) He I'CNCT – IN OP
If isolated: 10. (Aff) He I'CNCT – IN OP If TK P < 2200 @ MECO -60: 11. Shut dn MN ENG per MPS CMD/HYD/ELEC
Post ET SEP: 12. He I'CNCT(s) – GPC

Figure 1.4*. AESP Section headed* ***MPS He P (Pre MECO)****. Checklist for Main Engine helium supply systems malfunctions.*

dP/dT reading has not occurred until the vehicle is within 60 seconds of main engine cutoff time (MECO), the crew is instructed to shut down the engine manually (Step 2 in the checklist). The reference to "per MPS CMD/HYD/ELEC" is an instruction to go to a completely separate page in the FDF, where manual shutdown procedures are located. The double carat ">>" is a cue that if the crewmember does proceed to the engine shutdown section, he/she has finished with the current checklist and should not return.

The remaining instructions apply only if the helium problem has occurred earlier in flight (before MECO minus 60 seconds). Step 3 follows a second logical expression, and is relevant only if the "root cause" of the problem is a regulator failure, which the crew is instructed to determine by checking BFS GNC SYS SUM 1 for the presence of an up or down arrow beside either the A or B leg regulator pressures (abbreviated to "REG P" in the checklist and on BFS GNC SYS SUM 1). Suppose that Leg A REG P is indicating an off-nominal high value. The crewmember is instructed to close the "aff" (short for "affected" [Leg A]) isolation valve, thereby isolating Leg A and preventing additional helium from flowing through the failed regulator. If, instead, the "up" or "down" out-of-limits arrow was located beside the REG P B value, the correct response to the "aff He Isol – CL" instruction would be to close the Leg B isolation valve rather than Leg A.

Onward to "Otherwise". But suppose the abnormally high rate of helium depletion (dP/dT) is not due to a failed regulator. This situation corresponds to the "Otherwise" condition in the next line of the checklist. Having ruled out a regulator failure, the remaining possibility is a leak somewhere in the affected helium supply system, perhaps in the tank(s) themselves, in Leg A, or Leg B. The set of instructions following "Otherwise" is a sequence of procedures to try to isolate the leak to Leg A or Leg B. Leg A is targeted first; the crewmember is instructed to close ISOL A and check dP/dT to see if the reading returns to normal. If it does, the crewmember has succeeded in isolating the leak to Leg A. If closing ISOL A has no effect on dP/dT, the leak is not in Leg A and more steps must be taken. First, ISOL A has to be opened back up, and then ISOL B has to be closed. Again, the crewmember is prompted to check dP/dT to see if closing ISOL B brings the reading back to normal. If it does, the leak has been successfully isolated to

Leg B. If not, the leak is declared "nonisolatable", and the crewmember must navigate (skip) to the checklist section labeled "If nonisolatable".

Possibilities for crew error. Before describing the "If nonisolatable" procedures, we want to draw attention to an important human factors risk with these manual isolation procedures. The instructions are clearly ordered: If the leak has not been isolated to Leg A, open ISOL A back up, *and then* close ISOL B. Taking both steps in the correct order is safety critical; if the crewmember fails to open ISOL A, or closes ISOL B before opening ISOL A, he/she creates a situation where both ISOL A and B are closed simultaneously, choking off all helium to the main engine and causing an immediate engine shutdown. This example illustrates a more general point: any time a procedure calls upon the crew to manually reconfigure the operating mode of a safety-critical system, the procedure introduces the risk of a reconfiguration error, sometimes with potentially dire consequences to the mission. In this particular case, two distinct forms of FDF navigation error (skipping a procedure, or performing two procedures in the incorrect order) would result in an engine shutdown and mission abort. Many classes of mission aborts are extremely hazardous operational scenarios that are extensively trained in ground-based mission simulation. Whether the crew would survive an actual abort, however, is an untested and unknown question.

Replacing Lost Helium: The Backup (Pneumatic) System. For the moment, let's assume that the crew has succeeded in isolating the leak to either Leg A or Leg B. Even though the leak has been contained, and normal flow has been restored, the crewmember working the malfunction (the pilot) is not finished with the checklist. Note the section further down the checklist headed by the conditional expression, "If isolated". The first step in this section (Step #10 in the checklist) is written as "He I'CNCT – IN OP". To understand this instruction, just a little more information on the helium supply systems architecture is needed. All three helium supply systems are plumbed to a common manifold, which (among other capabilities) allows the crew to crossfeed helium from one engine's supply system to either (or both) of the others, should the need arise. In this particular case, however, the reason for this instruction is that the common manifold is pressurized with helium from a backup (fourth) supply system, known as the pneumatic system, one of whose functions is to provide supplemental helium to the main engines if needed. When the helium supply in one of the main engine supply systems has been depleted, as is the case when the system has experienced a leak, the instruction calls for the crewmember to toggle the affected supply system's common manifold interconnect valve to the "IN OP" (in open) position, allowing helium to flow from the common manifold into the affected Main Engine supply system. This step allows helium from the backup pneumatic system to supplement the supply remaining in the affected system's tanks.

Toggling the interconnect switch to "IN-OP" is also safety critical; without it, the helium supply may run out before the scheduled engine cut off time, again triggering an automatic engine shutdown and possible mission abort. But as depicted in the checklist, the actions associated with having successfully isolated the leak (the area headed by "If isolated") are segregated from the section containing the interconnect instruction, both spatially (the two boxes are physically separated by an interleaved subsection) and functionally (the two set of actions belong to different boxes). Crewmembers must thus overcome these "segregation cues" in order to

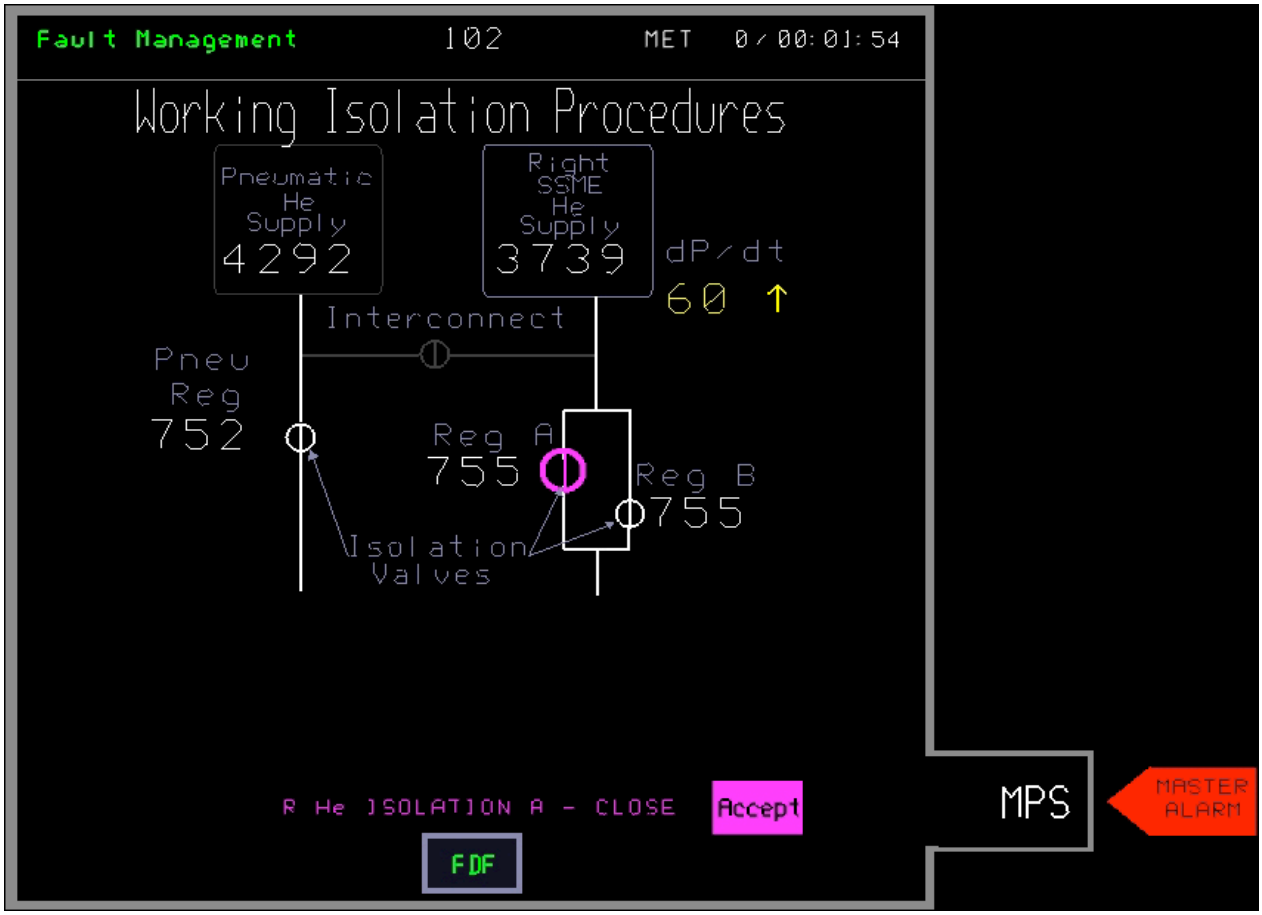


Figure 1.5. *Fault Management Display as it appears immediately following C&W annunciation of anomalous rate of change (dP/dT) in Right Engine helium supply system tank pressure. FAMSS has already identified the problem as one requiring isolation procedures.*

navigate successfully from one section to the other, and complete the full set of relevant procedures.

The final option: A "Nonisolatable" Leak. There is one last possible path through the checklist. Suppose the crew performs all the actions specified in Steps 4-6 (the isolation procedures) correctly, but the steps have no effect on the high dP/dT reading on BFS GNC SYS SUM 1. This situation corresponds to a "nonisolatable" helium leak, which would happen, for example, if the leak were in one of the helium supply tanks rather than in Leg A or Leg B. The appropriate set of instructions for this contingency is contained in the section headed by the conditional "If nonisolatable". The first instruction is to manually shut down the engine when the vehicle reaches 23,000 ft/sec of inertial velocity. The next instruction is to take the affected helium supply system's interconnect valve to the IN-OP (inlet-open) position when the tank pressure falls to 1150 psi. If the leak occurs early in the ascent phase, the "trigger" conditions that must be satisfied before the crew actually executes these procedures will not occur until as much as several minutes from when the crewmember first reads the instructions. Thus, the two procedures fall in the category of *deferred* procedures, qualitatively different from the other procedures in the section.

Summary. Having described the helium supply system checklist in considerable detail, we can now summarize the human factors problems that make paper checklist navigation so difficult. First, fast and efficient decoding of the instructions is compromised by the terse nature of the commands and numerous abbreviations. Second, the proliferation of logical conditionals forces the crew to constantly crosscheck FDF instructions against real-time sensed values on cockpit displays. Third, the outcome of these evaluations is crucial to accurate checklist navigation, since different outcomes typically engender different paths through the rest of the checklist. Fourth, checklist navigation is further complicated by the poor arrangement of the subsections; procedures that logically follow each other, and belong to the same navigation path, should be grouped together. Instead, they are often segregated by boundary delimiters and interpolated subsections. A major purpose of the FAMSS design is to eliminate or alleviate these difficulties with FDF navigation.

1.4.3 FAMSS Interfaces: The Fault Management Display

Figure 1.5 shows the primary FAMSS interface, incorporating many of the CAU design conventions, called the Fault Management Display. This display format integrates C&W system visual interfaces with an electronic version of off-nominal (FDF) checklists. We describe these features using an example Fault Management Display from one of our evaluation scenarios (see Section 2.6.2), an isolatable helium leak in the Right Engine helium supply system.

C&W features. The Fault Management Display serves as the primary crew interface with the C&W System. Consistent with ECW capabilities, the interface assumes a C&W system capable of making root-cause failure determinations and inhibiting superfluous C&W events generated by "downstream" events. Once the root cause is established, the malfunction is automatically associated with the system experiencing the problem, and the appropriate fault isolation and recovery procedures are retrieved and displayed automatically. Initially, the Fault Management Display appears as shown in Figure 1.5. In the lower right section of the display is a software Master Alarm light, a red rectangle whose left side points directly at a systems identifier tab. This replaces the hardware Master Alarm annunciator in today's shuttle cockpit, which annunciates the same alarm, in the same physical location, for all Class 1 and 2 C&W events. The FAMSS version is designed to make the Master Alarm more useful to the crew by using it to direct the crew's attention toward the tab that, in turn, identifies the system experiencing the malfunction. The crew extinguishes the light (and the accompanying auditory alert) by pressing the rectangle.

The procedures section. Having directed a crewmember's attention to the Fault Management Display, and to the system experiencing the malfunction, the next goal is to assist the crew with working through and executing the appropriate procedures. Armed with the assumed feature of computer access to all sensed parameters and automatic evaluation of FDF logical expressions, FAMSS determines the correct path through the FDF checklist. Only those procedures that are on the critical navigation path are displayed.

These procedures are depicted in the procedures box, the enclosed area to the left of the systems tab. Note that the left side of the tab is omitted, so that the tab opens on the display. This design feature tells the crew which malfunction is currently being worked, and whose procedures are currently represented in the procedures box. If there were multiple current unresolved

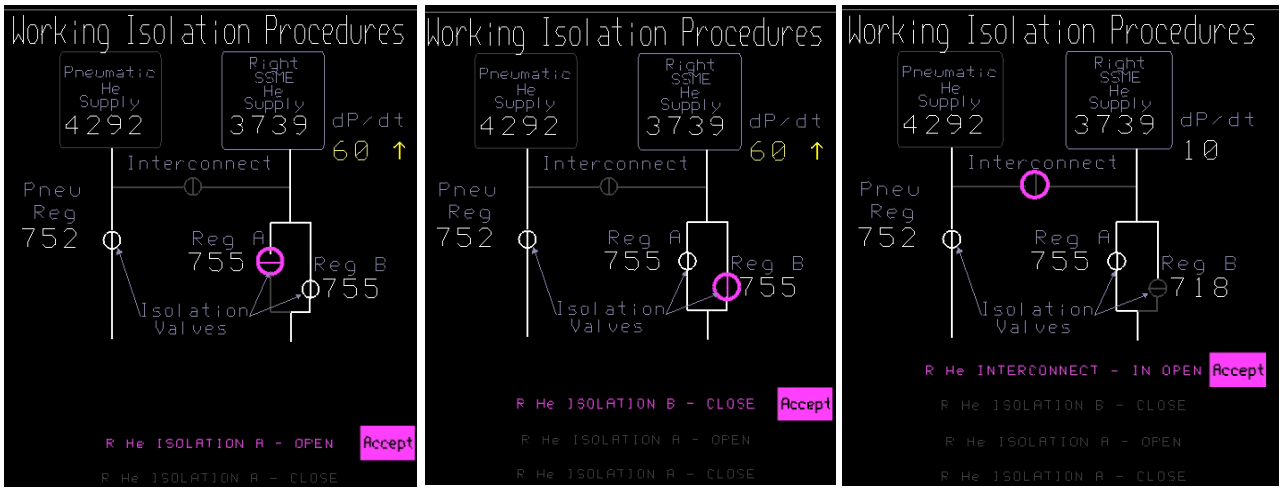
malfunctions, multiple systems tabs would appear along the right boundary. Crewmembers would select which malfunction to work by pressing the corresponding tab. That action would open up the left side of the selected tab and bring up that particular malfunction's procedures in the procedures box; the left boundary of all remaining tabs would be closed.

Currently Commanded Procedures. Figure 1.5 depicts a leak in the right engine helium supply system. The title of the procedures box, Working Isolation Procedures, reflects the fact that FAMSS has automatically evaluated the first and second logical expressions in the MPS He P section of the FDF (Figure 1.2), and automatically navigated to the "Otherwise" section. Thus, the first procedure displayed in the procedure box is Step #4 from the MPS He P section, "He ISOL A – CL," that is, close leg A helium isolation valve.

This instruction is represented in the fault management box in two distinct formats. In the lower section of the display, the instruction appears in as a magenta-colored line of text with content similar to Step #4 in the paper FDF (Figure 1.4). The only change is because FAMSS has identified which helium supply system is affected, so the "Aff" at the beginning of the procedure is replaced by "R" (for Right Engine), and there is enough display real estate to eliminate some abbreviations. Immediately beside the text instruction is a magenta "Accept" icon. If the crewmember accepts the automation's analysis and its procedural recommendation, he/she "gives permission" to the machine to take the action (in this case, Close ISOL A) by pressing the Accept icon. If the crewmember doesn't trust the automation, and wants to examine the logical expressions that have been evaluated in the course of navigating to this procedure, he/she can opt to bring up an electronic replica of the MPS He P FDF checklist by pressing the lower box with the embedded green "FDF."

Above the text area is a "zoomed-in" depiction of the relevant system schematic, a slightly modified version of the equivalent schematic on the CAU MPS SYS SUM display (Figure 1.3). Recall that the primary purpose of the schematic is to keep the crewmember in tight synchronization with the automation, by maximizing his or her awareness of the current operational mode of the system, what mode transition is being suggested, and what configuration will result. Thus, the "zoomed-in" section provides detailed information on the architectural components in the immediate architectural vicinity of the malfunction, including sensed parameters and current mode configurations. In the present example, the schematic is a slightly modified version of the right engine helium supply system schematic on the CAU MPS SUM display format (Figure 1.3).

The current procedure calls for closing ISOL A. This instruction is embedded within the schematic via redundant size and color-coding cues. Recall that on CAU display formats, circles with embedded line segments represent valves and their current configuration (OP [open] or CL [closed]). Since the crewmember has not yet given FAMSS permission to close ISOL A, the valve shows the current (open) configuration, with flow through Leg A. However, the symbol itself is physically larger the other valves on the display, and colored magenta, matching the color of the text and the "Accept" icon. The color and size coding indicate that FAMSS is asking permission to change the operating mode of the highlighted element. Since the valve is currently open, the instruction is for the valve to be closed; if the valve were closed, the coding would signal that FAMSS is recommending a change to open.



a b c

Figure 1.6. *Fault Management Display changes as FAMSS and crewmember work through helium leak isolation procedures. Figure 1.6(a) shows the Fault Management Display waiting for permission to complete Isol A OPEN procedure (Step 5 in MPS He P procedures section [Figure 1.4]); Figure 1.6(b) shows the Fault Management Display awaiting permission to complete Isol B CLOSE procedure (also part of Step 5); Figure 1.6(c) shows the Fault Management Display awaiting permission to complete right engine interconnect IN OPEN procedure (Step 10 in AESP MPS He P procedure section).*

Once this action is taken, the completed procedure is grayed out, and the next commanded procedure appears directly above it (Figure 1.6(a). The schematic changes to reflect the new system configuration (ISOL A closed) and depicts the color and size coding consistent with the new instruction. So, in Figure 1.6(b), the commanded procedure is to close ISOL B; in 1.6(c), ISOL B is now closed, dP/dT has dropped back to normal, and FAMSS is waiting for the operator to give permission to take the interconnect valve to IN OPEN (Step 10 in Figure 1.4).

Deferred Procedures. Recall that one of the navigation paths in the MPS He P procedures section leads to a nonisolatable helium leak, where the leak cannot be isolated to Leg A or Leg B. The correct response here is to, first, supplement the affected engine's helium supply by opening the interconnect valve, so helium can flow in to the system from the common manifold, and then shut the engine down manually. However, these are both *deferred* procedures; the interconnect valve is not supposed to be taken to the "IN-OPEN" position until the tank pressure drops below 1150 psi, and the engine is not supposed to be shut down until the vehicle reaches 23,000 feet per second (fps) of inertial velocity. Note that the MPS He P procedure section does not include the specific shutdown instruction. Rather, FAMSS automatically retrieves the MPS CMD/HYD/ELEC checklist and relevant sensed parameters to extract the appropriate instruction.

To distinguish deferred from immediate-action instructions, the instructions for deferred action are depicted in yellow or white rather than magenta (Figure 1.7). Utilizing FAMSS' assumed access to all relevant sensed parameters, the automation computes the difference (delta) between the helium tank pressure (currently 3410) and the tank pressure (1150) that will trigger the "IN OPEN" command. In the figure, the delta is 2260. To help prioritize the criticality of the tank

pressure reduction, FAMSS predicts the amount of time until that delta will drop to zero (given the current leak rate, and assuming that the rate holds constant). Similarly, FAMSS predicts the amount of time until the vehicle will reach 23K of inertial velocity. FAMSS then prioritizes the two procedures based on which event should occur first.

As illustrated in Figure 1.7, FAMSS has computed that the Center Engine helium tank pressure is diminishing at a rate that will take it to the critical tank pressure threshold (1150 psi) before the vehicle reaches the critical velocity threshold (23,000 fps). Accordingly, the text instructions for the Interconnect – IN OP action appears in yellow at the top of the procedures list, followed by the Center Main Engine Shutdown instruction in white. At the same time, the helium system supply schematic appears in the schematic section of the display, with the currently closed interconnect valve oversized and also colored yellow (indicating that it will eventually be commanded to the "Open" position, but not right away). Immediately beside the yellow text, the actual delta pressure value appears in digital form inside a rectangular countdown indicator. At a prespecified value for the delta pressure, a small black vertical slice appears along the left inside wall of the rectangle. The darkened region gradually expands to the right, eventually filling the rectangle, as the delta between the current pressure and the target pressure decreases. The countdown indicator for the deferred engine shutdown works in the same way, except the green digital value inside the countdown indicator is a direct temporal delta between current MET and the expected MET when the vehicle will reach 23,000 fps. A direct temporal countdown was deemed less desirable for the "Interconnect" instruction because the size of a leak is not necessarily stable.

The countdown indicators were designed to give crewmembers explicit visual information to more effectively time-share their nominal instrument and display scanning and fault management-related information processing. One of the hypothesis tested in our empirical evaluation of the FAMSS concept is that this and other Fault Management Display features would reduce cognitive tunneling on off-nominal information sources, and enable more nominal scanning during the deferred period.

The "deferred instructions" version of the Fault Management Display continues until the instruction converts to a real-time command. At that point, the magenta "Accept" icon replaces the countdown indicator, and the text line and instruction cues embedded in the schematic also turn magenta. No further changes are made to the display (except for updating parameter values) until a crewmember presses Accept. Then, the second (engine shutdown) instruction line and countdown indicator replaces the "Open Interconnect" instruction line, in yellow, and the helium supply system schematic is replaced with an engine shutdown schematic. This version remains on the screen until the engine reaches 23,000 fps of inertial velocity, at which time the countdown indicator is replaced with the magenta "Accept", and both text and Main Engine Shutdown valves also turn magenta. This situation is depicted in Figure 1.8.

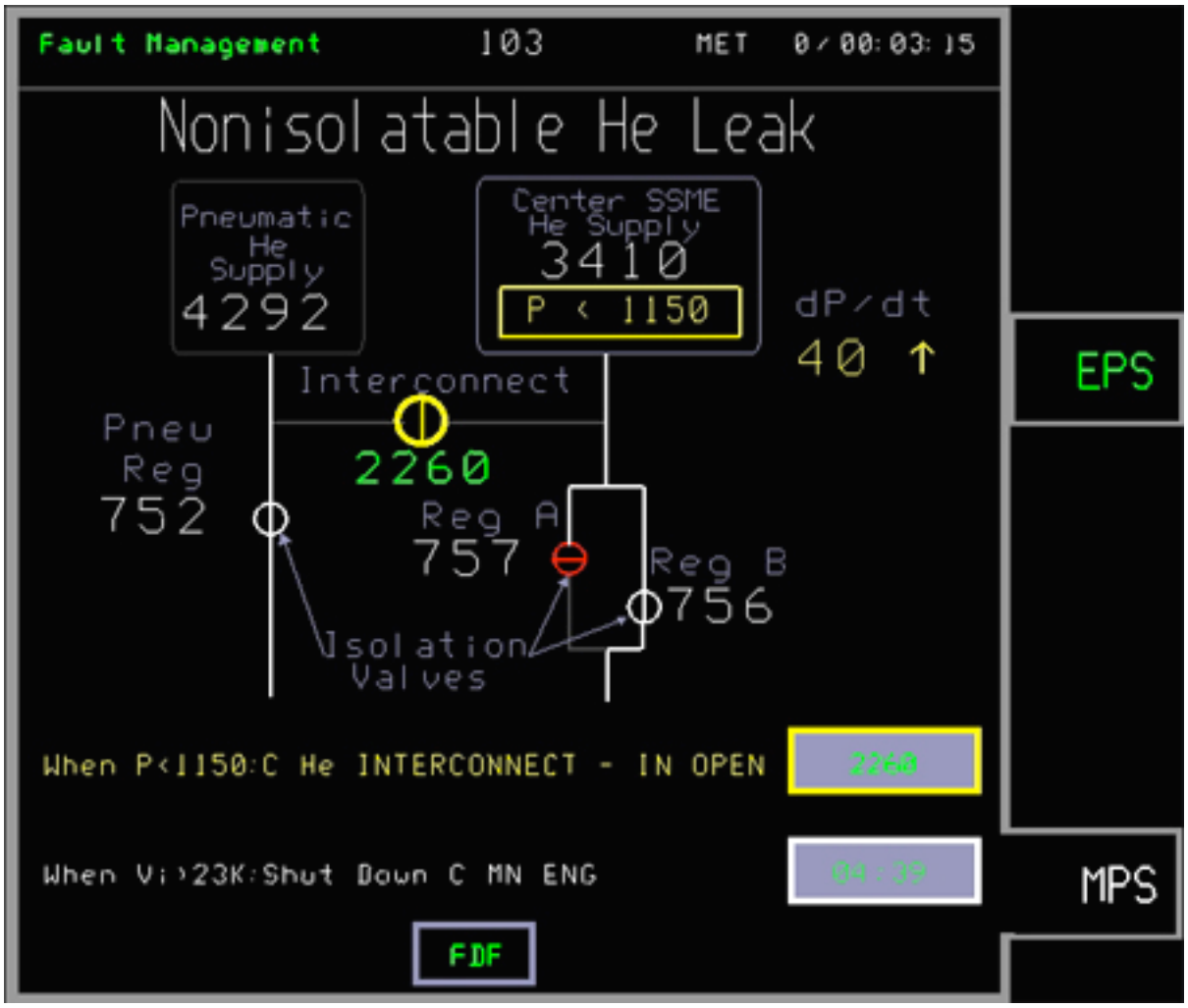


Figure 1.7*. Fault Management Display for deferred procedures as it appears immediately following MPS tab press for an nonisolatable leak in the Center Engine helium supply system. The two procedures are written in text form at the bottom of the display. Beside each procedure is a rectangular countdown indicator. In the case of the interconnect open (upper) procedure, the indicator contains a digital green pressure "delta" between the current Center Engine helium supply system tank pressure and the target pressure at which the interconnect will be commanded to OPEN. In the case of the shut down Center Engine (lower) procedure, the green digital is a direct temporal delta between current MET and predicted MET when the vehicle will reach 23,000 fps of inertial velocity. The graphical section above contains an oversize yellow interconnect valve symbol (.34" in diameter compared to .22" for the nonhighlighted valves), the target tank pressure inside the Center Engine Tank indicator, and the delta pressure in green below the interconnect valve symbol. See text for more details.*

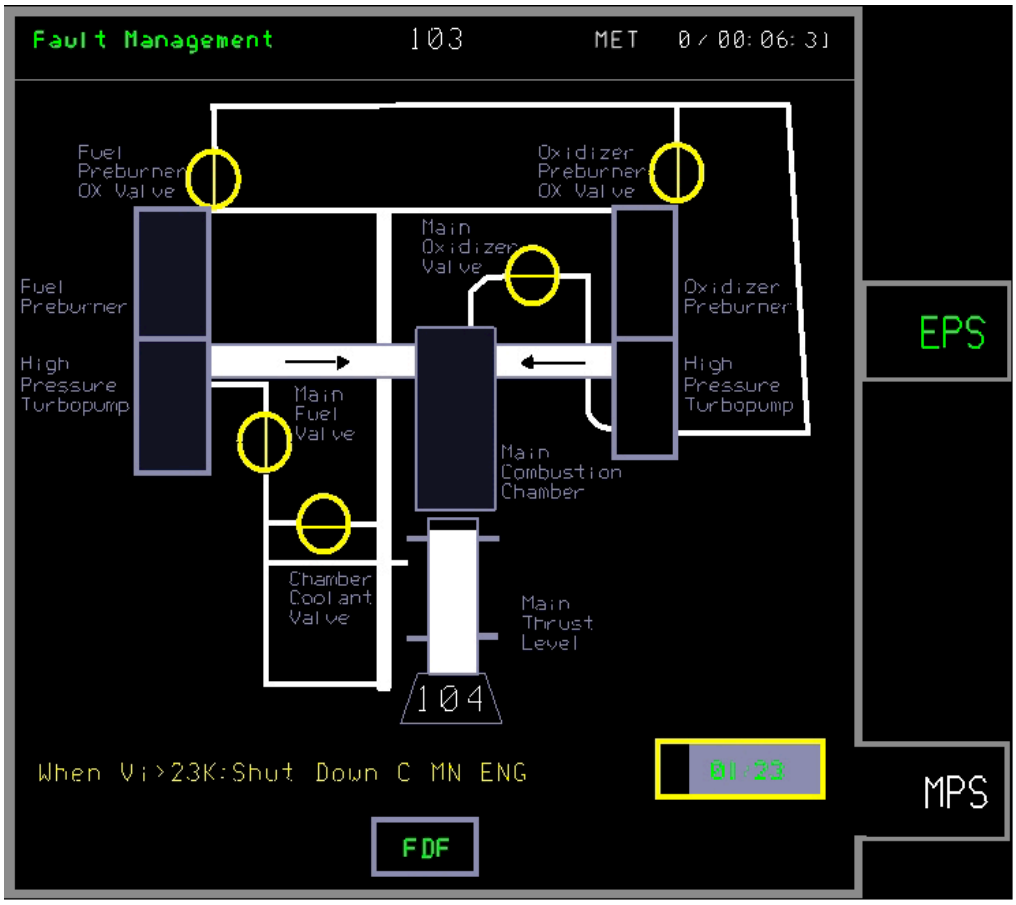


Figure 1.8. *Fault Management Display for deferred main engine shutdown procedure. The schematic section of the display contains a skeletal graphic of the main engine architecture complete with yellow main engine shutdown valves and thrust indicator from CAU MPS SUM. At the point in time shown in the figure, the vehicle is 1 min 23 sec short of reaching 23,000 fps inertial velocity. The black section of the countdown indicator has appeared and is enlarging to the right (replacing the grey). One min 23 sec from this time, the text and main engine shutdown valves turn magenta, and the countdown indicator is replaced by the magenta "accept" button.*

2 FAMSS Evaluation Methodology

We recently completed an extensive human-in-the loop empirical evaluation of FAMSS in part-task simulation. This section provide details about the evaluation methodology, including the selection of an appropriate "baseline" condition against which to measure FAMSS impacts, the simulation facility, participants, simulation tasks, and evaluation procedures. The results are described in Section 3.

2.1 Cockpit Conditions

Consistent with our assumption that CAU displays are the appropriate departure point for the development of next-generation spacecraft cockpits, and to more accurately quantify FAMSS-specific effects, participants performed a series of ascent-related malfunctions in two versions of the shuttle CAU cockpit. In the Baseline version, the key cockpit components were ascent-relevant CAU display formats, selected virtual switch panels, current C&W interfaces, and the Ascent-Entry Systems Procedures FDF document. In the FAMSS version, participants managed the same malfunctions in a version of the CAU cockpit modified to include the FAMSS cockpit interface (a CAU-style Fault Management Display format). This design allowed us to precisely capture and quantify FAMSS-related performance impacts, over and above any benefits to fault management associated with the CAU display formats themselves (which, as we noted, have already been quantified in both part-task and full-mission simulations). Henceforward, we will refer to the two versions of the CAU cockpit as the "Baseline" Condition and the "FAMSS" Condition.

2.2 Facility

The Intelligent Spacecraft Interface Systems (ISIS) laboratory is a reconfigurable facility with enough flexibility to simulate different suites of cockpit display formats and incorporate a variety of human performance measurement tools and evaluation techniques. The facility integrates six computers, twelve LCDs, eye tracking hardware, recording devices and a suite of software, including several core displays and the core flight model provided by NASA JSC.

The simulator relies on a highly distributed architecture. Six single-processor Intel Pentium 4 based personal computers, each with 512 MB of system memory and a dual-VGA Nvidea AGP card, are used to drive 12 touch-screen enabled liquid crystal displays (LCDs) that are all viewable by the experiment participant/crewmember. A professional mixer, along with a 6-speaker audio system, is also driven by the Intel-based computers to provide realistic sounds for engine, solid rocket booster (SRB) separation, and audible alerts. In turn, the Intel-based computers are synchronized with a flight model running on a dual-processor 250 Mhz SGI Octane over a Cisco gigabit ethernet switched network. On the experiment operator station (EOS) side of the equation, another Intel-based computer functions as a console to operate the simulator and control the data collection process, which collects real-time switch-throw and eye movement data. Yet another Intel-based system runs the eye-tracking equipment, (custom hardware designed by ISCAN, Inc. and Polhemus), with 9 video monitors attached for use in the setup and calibration process. The eye-tracking computer communicates with the EOS computer for data collection via a serial port running at a baud rate of 115,200 bits per second. Additionally, three Panasonic video disc recorders are used to collect audio and video of the runs as they unfold. Video recordings are overlaid with a millisecond timer as well as the simulator's mission elapsed time (MET) via video overlay processors (designed by Decade Engineering) that

Figure 2.1. *ISIS Cockpit in the Baseline (CAU) Configuration. Note the presence of the virtual hardware C&W matrix just above the operator's left shoulder. In the FAMSS Condition, the Fault Management Display format replaced this matrix. Prior to the occurrence of a malfunction (and throughout the nominal runs), this region was blank with the exception of a "All Systems Nominal" text message. The message returned once all malfunctions are resolved. The bright area in the top right-hand corner is an inserted still of a video recording based on the occupant's current fixation location, and was not present in the actual facility.*

are also driven via serial ports from the EOS computer for precise synchronization with the data collection process.

During a simulated mission, events unfold in units of simulator-driven MET, which are shown to participants in the top right hand corner of the CAU display formats in units of minutes and seconds. Because the networked processing scheme runs slightly slower than real (non-simulator) time, MET time is approximately 1.15 times slower than real (non-simulator) time. The results of our study are all presented in units of Mission-Elapsed (simulator) time, rather than real time. The only exception is in Appendix B, where the behavioral primitives associated with our human performance modeling work are expressed in units of real time.

As shown in Figure 2.1, the twelve LCDs surround the crewmember with display formats, gauges, and switches. The four LCDs directly in front of the cockpit seat represent seven of the nine cockpit Multifunction Display Units (MDUs). These LCD's are used to display any of the available CAU or FAMSS formats along with hardware panels such as the shuttle's hardware C&W matrix as well as several switches. The two overhead LCDs and five side LCD's are used to virtually represent a subset of the shuttle's gauges and switch panels. The remaining LCD substitutes for a shuttle keyboard. Because the LCDs are all touchscreens, crews can press the

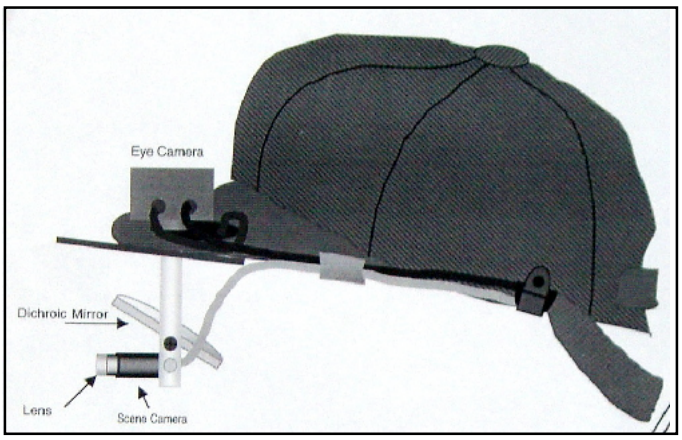


Figure 2.2. *ISCAN ETL-500 Cap-Mounted Eye Tracking Apparatus.*

simulated switches and keys to induce simulated changes in the cockpit (switch positions) and systems mode reconfigurations.

To determine what information participants are looking at on these displays, we use the ISCAN ETL-500 Cap-Mounted Eye Tracking Apparatus shown in Figure 2.2. The eye-tracker apparatus is an infrared, head-mounted tracker with a Polhemus head-tracker attached. This system is minimally burdensome (contained mostly on the visor of a lightweight cap) and computes eye positions at 60 Hz. The eye tracker works by shining an infrared LED (located in the visor) on the participant's eye and using a "hot mirror" (i.e., glass that primarily reflects only infrared light and is transparent to visible light) positioned in front of the eye to reflect an infrared image of the eye to an infrared camera (also located in the visor). The corresponding video image of the eye contains a bright spot on the cornea, corresponding to a reflected image of the infrared LED, and a dark image of the pupil. The tracker's image processing software computes the locations of the center of the corneal reflection and of the pupil, and then uses the distance between these combined with calibration information to determine the angle of the eye in the head. Eye angle information is then combined with head location and orientation information measured by the magnetic Polhemus head tracking system and used to estimate eye location and line of sight. Since the location of displays and other relevant objects is provided to the system during setup, the eye tracking system can extrapolate the point at which the line of sight intersects a plane (such as a group of displays) to estimate which display and approximately where on the display the person is looking. Because every person's eyes are unique, the eye tracking system is calibrated prior to each run. We describe this process in Section 2.6.4.1.

2.3 Participants

Fourteen recently retired commercial airline pilots, with an average of approximately 16,000 flight hours on various aircraft, participated in our evaluation. Highly experienced airline pilots were targeted because of their familiarity with complex mechanical systems, flight dynamics and cockpit scanning techniques. Prior to being selected for the ISIS experiments, none of the pilots had any operational experience in a spacecraft cockpit. This eliminated the possibility for specific-cockpit bias and allowed us to control the type, amount and schedule of training. The training regimens for previous studies in the ISIS lab and for the study described in this report are described in the next subsection.

2.4 Training and Testing

2.4.1 History

All fourteen participants took part in two previous ISIS studies, each with a training and testing regimen that bears directly on the level of expertise they brought to the present evaluation. The initial study, performed in 2002, investigated whether and how much CAU display formats

improved malfunction handling capabilities, compared to the MEDS display formats in use today, under relatively low-workload (single malfunction) conditions. The fourteen pilots were divided into two equal groups. Both groups separately participated in a 5-day curriculum of classroom lessons covering basic shuttle systems, ascent-related displays, display navigation (i.e., keyboard) functions, nominal display monitoring requirements during ascent, and procedures for working several possible malfunctions. For each potential malfunction, participants were instructed in the proper procedures for resolving it, and where to access these procedures in the AESP FDF. The only difference was that one group was trained in the MEDS cockpit, whereas the other group was trained in the CAU cockpit. Following the classroom training, each pilot was given a two-hour familiarization and practice session in his respective cockpit simulator.

Each pilot then completed several ascent runs, some nominal and some off nominal. The off-nominal runs contained either a regulator failure in the helium supply subsystem for one of the shuttle's three main engines, an ullage pressure problem in the external hydrogen tank, a failure of one of the five onboard GPCs, or a failure in the vehicle's flash evaporator system that cools the freon loops (critical components of the Environmental Control and Life Support System) following Solid Rocket Booster separation. Approximately four months after completing the single-malfunction study, each group of participants returned for a one-day refresher course (in either the MEDS or CAU cockpit displays) covering the basic shuttle systems, nominal monitoring tasks during ascent, and resolution procedures for the targeted set of systems malfunctions that could be simulated in our facility. As part of this refresher, each pilot was given a one-hour familiarization session in his respective cockpit. He then completed four test runs, two nominal and two off nominal. In contrast to the first study, the pilots were tasked with handling three malfunctions during the off-nominal runs. The set of possible malfunctions was the same as in the earlier single-malfunction study.

In order to have both groups of pilots gain equal familiarity with the CAU cockpit, the MEDS-trained group returned a year after the multiple-malfunction study for conversion training. The group attended three days of classroom training on the CAU cockpit, and then completed the multiple-malfunction study once again, this time in the CAU cockpit. Thus, prior to the current study, all fourteen pilots received nearly equivalent training in the ISIS version of the CAU cockpit.

More details and results of these studies are available in Hayashi, Huemer, Renema, Elkins, McCandless, & McCann (2005); Hayashi, et al., 2005; Huemer, Hayashi, Renema, Elkins, McCandless, & McCann, 2005; Huemer, et al., 2005; and Matessa & Remington (2005a,b).

2.4.2 Procedures for the Current Study

As part of their training for the current study, participants attended one day of classroom training that reviewed ascent-related CAU displays and previously experienced malfunctions. They also received an overview of the Electrical Power System (EPS), FAMSS display formats and user interfaces, and several new malfunction scenarios, including an aft power controller 4 (APC4) electrical subbus failure and associated cockpit signatures. The malfunction training was considerably more sophisticated than previous training, including an explanation of how the

APC4 sub-bus failure affects other systems, and how to interpret the multiple simultaneous C&W events that accompany the APC4 failure as symptoms of the underlining EPS failure.

As with the previous training regimens, classroom instruction was particularly focused on malfunctions of the main engine helium supply systems. The pilots were provided refresher training on the helium supply systems architecture, how to infer current system operating mode on MPS SUM and the FAMSS Fault Management Display schematics, how malfunctions affect the workings of the helium supply systems, and how to recognize malfunctions by their signatures on the CAU System Summary Displays. They received specific instruction on how to navigate the FDF MPS He P checklist for both isolatable and nonisolatable leaks. The FAMSS display of the helium supply system was covered in considerable detail. Further, throughout the training, pilots could ask questions to get clarification on any aspect they did not understand.

In addition to the technical details of the targeted systems (such as the helium supply system), the following four general concepts were emphasized:

1. Scan.
    - The pilots were instructed about the six required checks that must be completed shortly after liftoff.
    - For nominal situations, they were taught the PAHUEE scan, described in the next section, and asked to perform it at least every 30 seconds.
    - For off-nominal situations, they were warned about attention tunneling and reminded not to forget to conduct the larger scan periodically looking for additional malfunctions.
2. Crosscheck.
    - The pilots were encouraged to confirm that switch movements result in the expected mode transitions.
    - In the FAMSS Condition, they were encouraged to cross-check with other instruments before accepting an action.
    - In the FAMSS Condition, they were taught to confirm all switch positions after accepting an action.
3. Toggle Alternate displays.
    - In some situations, two displays may share one LCD. The pilots were reminded to check both displays in these situations.
4. Use the FDF as a verification list, not a work list.
    - Airline pilots are trained to use checklists as work lists; they first read the step on the checklist and then perform the action. Astronauts tend to use the checklists more as verification lists; when confident that they know the procedure, they perform the steps from memory and then verify by reference to the checklist that they performed all the required actions.

2.4.3 Same-Day Training and Testing Description

Approximately two months after their classroom training, all fourteen participants returned for two days of further in-simulator training and data acquisition. To determine the impacts of the FAMSS concept on participants' malfunction handling capabilities, we employed a two-day, four-run per day study. As explained above, each participant was trained and participated in data collection in both cockpit conditions (one condition per day). On each day, runs 2 and 4 were

nominal, with no simulated system malfunctions, and were used to collect baseline measures of nominal scanning behavior and to slightly "soften" participants' expectancies of encountering a malfunction. Runs 1 and 3 were off nominal. One off-nominal run contained a single malfunction – an isolatable helium leak in the right-engine helium supply system. The other contained three distinct malfunctions – an APC4 subbus failure in the EPS, a loss of processing synchronization (fail to synch) in one of the five onboard GPCs, and a nonisolatable helium leak in the center-engine helium supply system. Participants were not informed whether the upcoming run would be nominal or off nominal.

To address possible practice effects, we counterbalanced both Cockpit Condition (Baseline versus FAMSS) and Scenario Complexity (Single Malfunction versus Multiple Malfunction). Seven participants were assigned to Baseline Condition training and data collection on Day 1 and to FAMSS training and data collection on Day 2. For the remaining participants, this assignment was reversed. Moreover, on Day 1 a randomly selected seven participants received the single malfunction scenario on the first data collection run and the multiple-malfunction scenario on the third data collection run. On Day 2, this ordering was reversed. The remaining seven participants received the multiple-malfunction scenario on the third data collection run and the single malfunction run on the first data collection run on Day 1, and the reverse assignment on Day 2.

Participants received two consecutive mornings (approximately 2.5 hours duration) of same-day in-simulator training, one morning in the Baseline Condition and one in the FAMSS Condition. Then, in the afternoon, they completed the 4 data collection runs in the same cockpit condition (Baseline or FAMSS) as the morning training sessions.

Participants were trained and tested in pairs. The same-day (morning) training sessions were divided into two sections, each with six training runs. In the first section, one participant sat in the seat and acted as the shuttle commander; the other participant acted as a cockpit observer. Accompanying the two participants was an experimenter who acted as a subject matter expert (SME) "consultant," explaining important aspects of the malfunctions, controlling progress through the training runs, and coaching the participant through the appropriate procedures. In the second section, the "commander" and "observer" exchanged roles, and the six training runs were repeated.

The first training run contained a Leg A regulator malfunction in the Center Engine helium supply system. The second run included a low ullage pressure problem in the External Tank, the third contained a high Auxiliary Power Unit (APU) oil temperature problem, and the fourth contained a GPC fail to synch malfunction. The penultimate training run included an APC4 subbus malfunction. As soon as the malfunction occurred, the simulator was frozen, the cockpit signatures of the "downstream" impacts were pointed out and explained, and the meaning and importance of the "Do not Isolate" instruction in the AESP FDF were emphasized. This was followed by the final training run, which included a nonisolatable leak in the Center Engine Helium Supply system.

As each malfunction occurred, the "active" (seated) participant attempted to work the problem while the other participant and the experimenter SME observed his progress. Where it was

deemed appropriate, the SME talked the participant through the correct procedures and FDF navigation procedures. Note that in the case of the FAMSS condition, each "practice" malfunction had its own customized Fault Management Display. Where possible, the schematics section of the display provided embedded cues to the relevant procedures.

2.5 Data Collection Runs

Following a lunch break, each pilot participated in the data collection phase of the study, which consisted of four runs: two nominal and two off nominal. The following subsection describes the displays and information acquisition activities involved in the nominal runs, and describes a scan pattern developed to assist our non-astronaut participants in performing these activities. Then, subsection 2.5.2 describes the information acquisition activities and performance elements on the off-nominal runs.

2.5.1 Nominal Runs

Although the nominal runs contained no malfunctions, they did place information acquisition and information processing requirements on participants. On off-nominal runs, fault management activities are "superimposed" on these nominal requirements, raising interesting issues of time-sharing and attentional allocation that, in many cases, can only be addressed through eye movement analyses. The following section provides an overview of these activities.

2.5.1.1 Display Overview

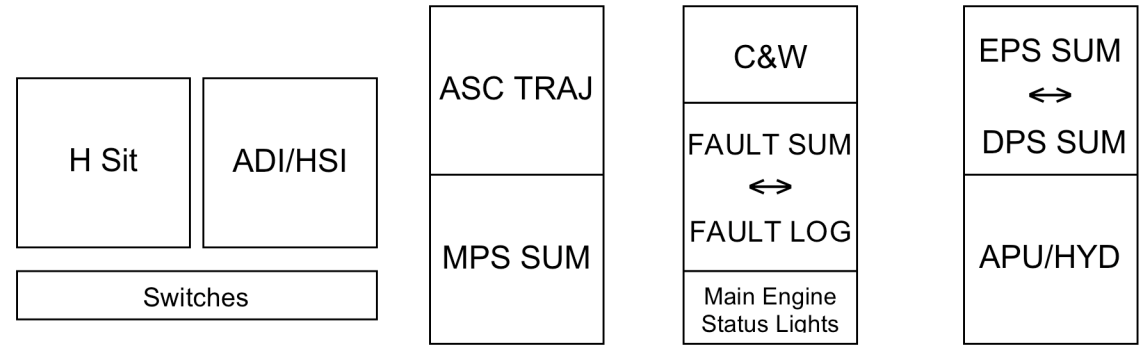
Figure 2.3 illustrates the arrangements of the forward displays in the Baseline and FAMSS conditions. Appendix A shows colorized screen shots of the major displays and indicates display regions containing information that should be monitored during the nominal runs.

2.5.1.2 Nominal Ascent Checklist Item Monitoring

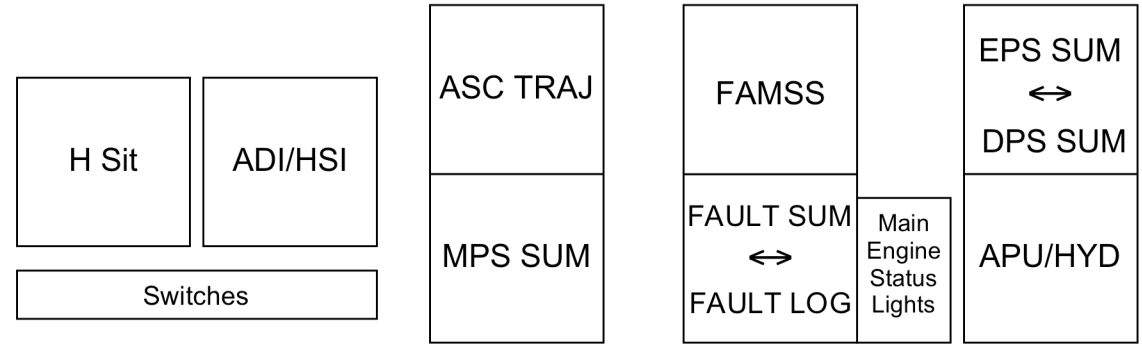
Table 2.1 lists the monitoring tasks required in the nominal ascent operations checklist. These monitoring tasks must be performed at specific times in order to certify that the vehicle systems are performing various pre-programmed ascent operations correctly. We call this type of monitoring the *ascent checklist item monitoring*. Note that the only physical action required during a nominal ascent is to take the ADI ATTITUDE switch below the ADI/HSI display to the LVLH position around 0:07 mission elapsed time (MET). Thus, during most of the ascent phase the operator is primarily monitoring highly automated vehicle systems and flight operations.

2.5.1.3 Nominal Scanning

When there was no specific event to monitor, participants were instructed to scan the front and overhead displays to remain appraised of vehicle trajectory, current abort options, and key operating parameters. We call this type of generic sequential acquisition of information from cockpit instruments and displays *nominal scanning*.



(a) Baseline Cockpit



(b) FAMSS Cockpit

Figure 2.3. *Forward Display Arrangements in the Baseline and FAMSS Conditions. (The double sided arrow means that the operator can toggle these displays).*

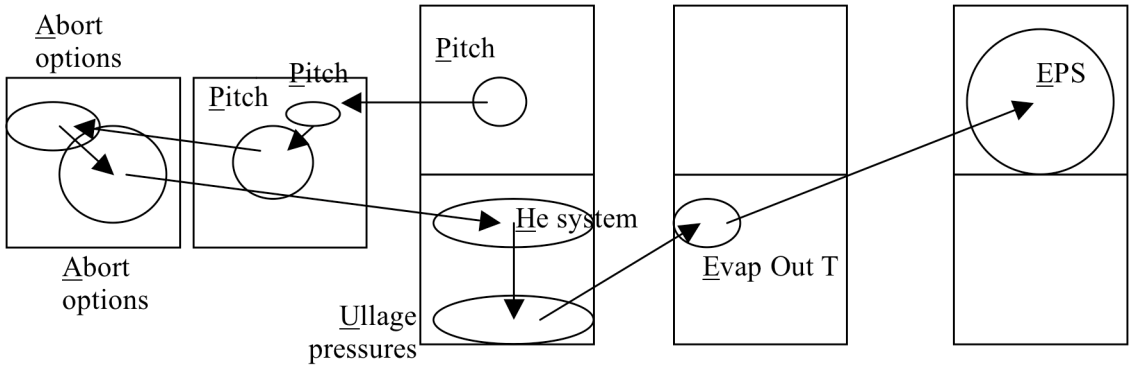


Figure 2.4. *"PAHUEE" scan (Baseline and FAMSS Conditions).*

Table 2.1. Nominal ascent checklist item monitoring tasks

Approximate Mission Elapsed Time (MET) [min:sec]	Ascent Phase Operational Event Names	Monitoring Tasks
0:07	Roll Program	Monitor that vehicle rolls to heads-down ascent attitude on ADI/HSI. Select LVLH (Local Vertical Local Horizontal) on ADI ATTITUDE switch located below H Sit.
0:30-1:00	Thrust Bucket	Monitor on MPS SUM or ASC TRAJ that main engines' throttling reduces to 67% in order to limit maximum flight dynamic pressure, then 30 seconds later, comes back to 104%.
2:00	Solid Rocket Booster (SRB) Separation	(i) Monitor on ASC TRAJ that pressure level inside SRB chamber (Pc) falls below 50 psi, which triggers SRB separation. (ii) Verify that the top part of each display indicates that Major Mode (MM) changes from 102 to 103. (iii) Check on the ASC TRAJ that Time to MECO (TMECO) values computed by the PASS and BFS converge.
3:00		Verify the flash evaporator temperature is < 60°F and decreasing on FAULT SUM and Evap Out T overhead meter.
5:40	Roll to Heads Up	Monitor on the ADI/HSI that vehicle rolls to heads-up attitude.
8:30	Main Engine Cutoff (MECO)	(i) Check the MPS thrust level drops to 0% on the MPS SUM. (ii) Check that MAIN ENGINE STATUS lights illuminate red. (iii) Monitor Cutoff bug on ASC TRAJ indicates MECO velocity.

To facilitate as comprehensive a nominal instrument scan as possible, participants were trained to perform a "PAHUEE" scan, which is to monitor (1) Pitch values on the ASC TRAJ and the ADI/HSI, (2) Abort option on the CAU horizontal situation ("H Sit") display format, (3) Helium system status on MPS SUM, (4) Ullage pressures readings from the external tank on MPS SUM, (5) Evap Out T readings on Fault Sum, and (6) EPS status on EPS SUM, in this order. When executed, the PAHUEE scan covers most of the major displays from left to right (see Figure 2.4). For further clarification of these information sources, Appendix A contains colorized examples of each nominal CAU display format.

2.5.2 Off-Nominal Runs

2.5.2.1 Overview

As already noted, one of the two off-nominal runs contained a simulated leak in Leg B of the Right Engine helium supply system. The other contained three distinct malfunctions: a failure in

the EPS APC4 subbus, a loss of processing synchronization in GPC 4 (one of the five onboard GPCs), and a nonisolatable helium leak in the Center Engine helium supply system.

Each malfunction required a clearly specified set of cockpit information processing or manual operations. Participants had to first extinguish an auditory/visual alarm, and then identify the root cause of the malfunction by deciphering the C&W fault message(s) on the CAU FAULT SUM Display and studying relevant cockpit displays (including the FAMSS Fault Management Display, when available). Once the malfunction was identified, participants had to locate and carry out the instructions in the relevant section of the AESP FDF, or on the FAMSS Fault Management Display. Three of the four malfunctions called for changes to the operational mode of the affected system, and so carried a requirement to locate cockpit switches and toggle them to new positions. Whether these switch throws were performed manually or by machine also depended on cockpit condition. In the following two subsections, we describe the malfunctions and how they were handled in each condition.

2.5.2.2 Single-Malfunction Run

2.5.2.2.1 Baseline Condition

Single-malfunction runs were nominal until 1:50 MET, ten seconds prior to solid rocket booster separation. At that point, an isolatable leak in the Right Engine helium supply system was simulated. The C&W system annunciated the event with a rapidly alternating tone, illuminated (red) Master Alarm light, and a fault message on Fault SUM reading "MPS He P." Meanwhile, on MPS SUM, the dP/dT numeric value for the right engine helium supply system started to read more than 20 psi, and both the numeric value and adjacent up arrow were colored red.

To determine the cause of the malfunction, the participant must flip through the AESP, locate the MPS section, and then find the procedures headed by MPS He P. Assuming navigation and procedure execution proceed correctly, as described in Section 1.4.2, the participants first attempt to isolate the leak by closing ISOL A and checking MPS SUM to determine whether the dP/dT reading has returned to normal. In fact, the simulated leak was in Leg B, so the abnormally high dP/dT reading was unaffected by closing ISOL A. The next step was to reopen ISOL A and close ISOL B. This action did cause dP/dT to drop back to normal, signaling a successful isolation effort. The final instruction, 10 in the checklist, called for one last switch throw to connect the affected engine's helium supply system to the common manifold, making supplemental helium available from the backup pneumatic system.

We noted earlier that designing redundancies into spacecraft systems is a critical means of reducing the risk posed by systems malfunctions. However, the downside of architectural redundancy is that redundancy management typically requires crew actions to change the operating mode of the affected system, introducing the risk of human error. In the case of this helium leak, the most critical risk is that the operator could inadvertently close ISOL A and then close ISOL B before reopening ISOL A, thus choking off all helium supply to the engine and causing an immediate abort.

2.5.2.2.2 FAMSS Condition

In the FAMSS condition, participants worked the identical helium leak malfunction with the identical display suite, except that the C&W annunciator matrix was replaced by the Fault Management Display described in Section 1.4.1 (FAMSS Interfaces; see also Figure 1.5). In addition, the visual Master Alarm light was disabled and replaced by the more localized and informative software visual alarm embedded in the Fault Management Display.

Determining the cause of the malfunction in the FAMSS cockpit is greatly simplified. Specifically, based on the sensor data, FAMSS automatically determines that the problem is a potentially isolatable helium leak. It automatically navigates to the correct procedure in the FDF and displays the first applicable instruction to the participant. The participant then needs to "Accept" the presented instructions, all along verifying that the instruction makes sense in the given context by cross checking, as previously taught. Moreover, FAMSS automates switch throws, eliminating the need to physically locate switches. Additional details of the Fault Management Display and Operator-FAMSS interactions in conjunction with this malfunction were described in Section 1.4.3 (FAMSS Interfaces).

2.5.2.3 Multiple-Malfunction Run

2.5.2.3.1 Baseline Condition

While the isolatable helium leak on the single malfunction run was obviously not without its complications (and, as we shall see, not all participants completed all FDF procedures correctly), the procedures were logically straightforward and self-contained (i.e., they were not conditional on cross-system impact assessments). The multiple-malfunction run was considerably more complicated. The first of three malfunctions, a failure of subbus APC4 in the EPS distribution network, occurred 30 sec after liftoff, at the same time as the main engines were entering the "thrust bucket" (Table 2.1). The most direct indication of this failure in the nominal ISIS display configuration is a red subbus indicator on the CAU FAULT SUM display. The EPS SUM display has a much larger segment of the power distribution section that is also red. However, the default display for this LCD is DPS SUM, not EPS SUM, so accessing the APC4 failure indication required display navigation (pressing the EPS toggle soft key). Recognizing the root cause of this failure was also complicated by the fact that an APC4 failure has several distinct cross-system impacts and cockpit indications. On the Fault Summary Display, the C&W system generates two fault messages, "APU 3 SPD LOW" and "MPS LH2/LO2 ULL P," neither of which provides a direct indication of the root cause failure. The APU failure message corresponds to a failure signature on the APU/HYD SUM display, in the form of a reference to a Remote Power Control distribution failure that appears in red, and the APU 3 percentage of normal RPM value shows a value of zero in yellow. The ullage pressure message corresponds to a low Center Engine ullage pressure reading on MPS SUM that appears in yellow with a down arrow beside it (illustrated in Figure 1.3). The danger here, of course, is that the combination of fault messages and off-nominal cockpit signatures would capture the operator's attention to the point where he/she fails to associate the signatures with the appropriate root cause, and starts to work one of the daughter problems.

The third and potentially most serious consequence of the APC4 failure did not generate a direct caution and warning event (alarm and fault message). However, on the shuttle, all of the main engine helium isolation valves are electrically actuated to the "Open" position, and the APC

subbuses supply power to the valves. When APC4 fails, power is removed from the Center Engine ISOL A valve, causing it to fail closed. Although this failure does not generate a C&W event, CAU display logic automatically turned the ISOL A Valve symbol on MPS SUM red, showed the valve as closed, and converted the Leg A feedline below the A indicator from white (signaling "flow") to dark gray ("no flow"). These signatures are illustrated in Figure 1.3.

For now, let's suppose that the crewmember assimilates all of these cockpit signatures, and correctly recognizes them as being due to an APC4 subbus failure. As with other malfunctions, the correct next step is to locate the appropriate procedures for an APC4 failure in the Data Processing System (DPS) section of the FDF. That section, titled "SUBBUS [APC4(5,6) or ALC1(2,3)]," does not contain any actual actions to fix the problem. Rather, the entire section is composed of a blunt command: Do not isolate MPS He C(L,R). The reason is straightforward: if the affected helium supply system were to experience a problem later on in flight (such as abnormally high dP/dT) crewmembers should *not* attempt to isolate the leak to one leg or another pursuant to the normal path through the FDF because, with ISOL A already failed closed, only Leg B has flow. Performing the standard set of malfunction procedures would result in both isolation valves being closed, leading to immediate engine shutdown and mission abort. Thus, the instruction is tantamount to telling the crew that if the affected helium supply system experiences a problem (leak or regulator failure) later on, the problem is to be treated as nonisolatable.

And indeed, at MET 3:00, we did introduce a leak in the affected (Center) Engine helium supply system. The correct navigation path through the FDF checklist therefore began with the "If nonisolatable" section (Figure 1.4). As described in Section 1.4.3, the crewmember has two actions to perform, the second one being a manual shutdown of the Center Engine when the vehicle reaches 23,000 fps of inertial velocity.

The situation may seem straightforward, but it represents something of a human factors time bomb. Just 20 sec after the initial APC4 failure, an unrelated malfunction occurs in the shuttle's data processing system. Specifically, the fourth of the four onboard general purpose computers (GPC's) that redundantly and simultaneously process primary flight software commands (collectively known as the "Primary Avionics System Software [PASS] set), stops processing in tight synchronization with the other three PASS computers. The data processing system also contains a fifth GPC, loaded with "backup" flight control and systems management software that is not part of the "PASS" set. The GPC4 fail to synch failure generates a C&W alarm and two fault messages on the CAU FAULT SUM Display, one from the PASS set, the other from the backup flight system (BFS) computer. Again, the correct process is to identify the correct cause of the fault messages, locate the corresponding main procedure title ("PASS GPC FTS") in the data processing system section of the AESP, and navigate through the fail to synch checklist.

Assuming the checklist was navigated correctly, three switch throws were called for, the first to take Flight Control System Switch #4 (located beneath the keyboard in the ISIS lab) to the closed position, and the remaining two being consecutive toggles of a three-position GPC 4 Mode Control Switch on Overhead Panel 2. The first toggle takes the mode control switch from run to standby, which the crewmember should hold for three seconds to allow the data processing system time to save a copy of the computer's software, and the second toggle takes the switch to

Halt. This action stopped further processing by the malfunctioning PASS computer and prevented it from issuing errant flight commands to the main engines.

The GPC fail to synch failure occurs at MET 50 sec, only 20 sec after the APC4 failure, and just about half way through the first stage of powered flight, when the Solid Rocket Boosters are still firing. As we noted, there was a good chance that participants would not have finished processing the APC4 problem at that point, in which case they had to time-share fault management activities for two malfunctions. A little later, at MET 2:00 minutes, the Solids Rocket Boosters separated from the vehicle. Both PASS and BFS flight software went "closed-loop" for attitude control at that point, and the two software systems independently computed when Main Engine Cut-Off (MECO) should occur. Both the PASS and BFS flight software should have agreed on this time, and one of the most important nominal checks for the crew was to verify that they did so. Thus, even if the GPC fail-to-synch malfunction was resolved by that point, the operator was likely to stay fairly busy. It was not until MET 3:00 that we introduced the third malfunction, a leak in the Center Engine helium supply system. This is, of course, the same system that was already operating in an off-nominal mode because ISOL A was failed closed. Even with all the intervening activities and diversions, the participant had to remember that this was a nonisolatable situation, and navigate directly to the "if nonisolatable" section of the FDF checklist.

Recall from Section 1.4.3 that the nonisolatable condition involves two deferred procedures: toggle the center engine interconnect valve to the "IN-Open" position when the affected system's tank pressure bleeds down to 1150 psi, and shutting the engine down when the vehicle reaches 23,000 fps of inertial velocity. These conditions were not satisfied until several minutes after the helium leak occurred, and it was up to the operator to monitor both the center engine helium tank pressure (on MPS SYS SUM) and the vehicle velocity on the velocity tape (located to the left of the ADI indicator on the ADI/HSI display) to know when to carry out these actions.

Of course, this discussion assumes that the operator remembered that the system is in a "do not isolate" mode, due to ISOL A being failed closed. Thus, the instruction associated with the APC4 failure - do not isolate MPS He - is essentially a prospective memory instruction that has to be remembered at the time of the actual helium malfunction. As we have seen, our multiple malfunction run contained a great deal of activity, including an entirely independent malfunction, between the time when the APC4 failure implications are digested, and when the helium system malfunction actually occurred. Prospective memory is one of the most fragile and easily disrupted forms of human memory under high-workload conditions, like those on a multiple-malfunction mission in a spacecraft cockpit. Thus, responding to all three malfunctions in the appropriate manner presented quite a challenge to our participants.

2.5.2.3.2 FAMSS Condition

The multiple malfunction run in the FAMSS condition included the same malfunctions, at the same times, as in the Baseline condition, but in a version of the CAU cockpit modified to include a fully interactive FAMSS Fault Management Display. A static screen shot of the Fault Management Display that accompanied the APC4 failure is reproduced in Figure 2.5; that accompanied the first procedure for the GPC 4 fail to synch problem, in Figure 2.6; and that accompanied the nonisolatable helium leak in the Center Main Engine, in Figure 1.7 (the

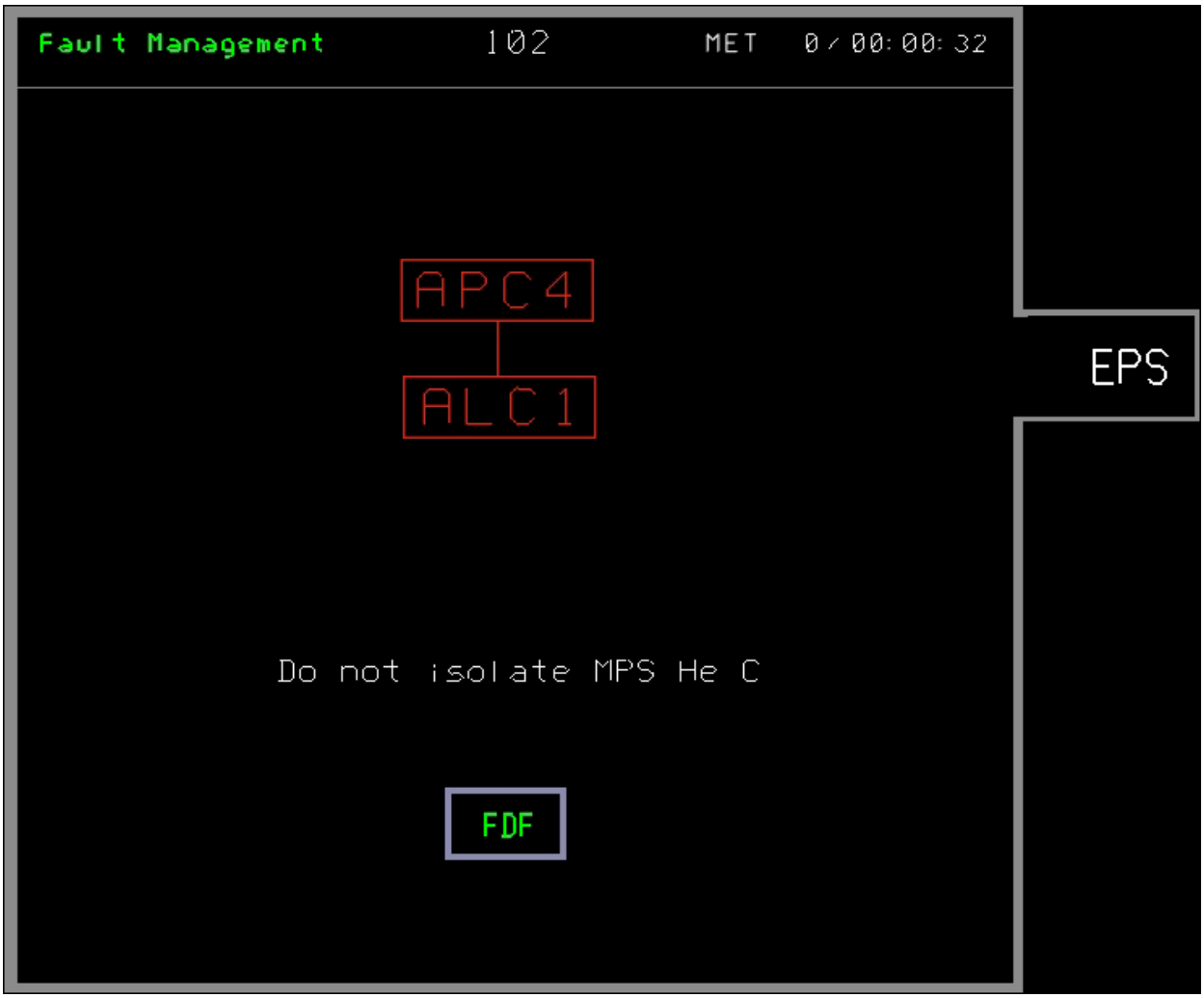


Figure 2.5. *Fault Management Display for EPS APC4 subbus failure. The text section instructs the operator not to attempt to isolate the Center Engine helium supply system in the event of a leak. The graphics section replicates the depiction of the failed sections of the distribution assembly on CAU EPS SUM.*

"deferred procedures" example from Section 1.4.3). There are several noteworthy features of these display formats. First, unlike the helium system malfunctions, neither the APC4 subbus or the GPC fail to synch malfunctions lent themselves in any straightforward manner to embedding procedural cues inside a system schematic. Thus, in the schematic section of the Fault Management Display, we attempted to consolidate the most relevant information concerning the malfunction from the various information sources spread across the cockpit. So, for the GPC fail to synch problem, the left side of the schematic section contains a "port" of the GPC matrix on panel overhead 2 (O2); the column beside the matrix is a portion of the CAU DPS SUM display showing the affected computer and a section of the data bus assigned to it. Similarly, the "schematics" section of the APC4 Fault Management Display shows the red APC4 section from the DPS SUM display. The goal here was to consolidate as much information as possible about the fault and its management activities on a single reference display, just as the CAU FAULT SUM display consolidates "big picture" information about the health of all significant spacecraft systems.

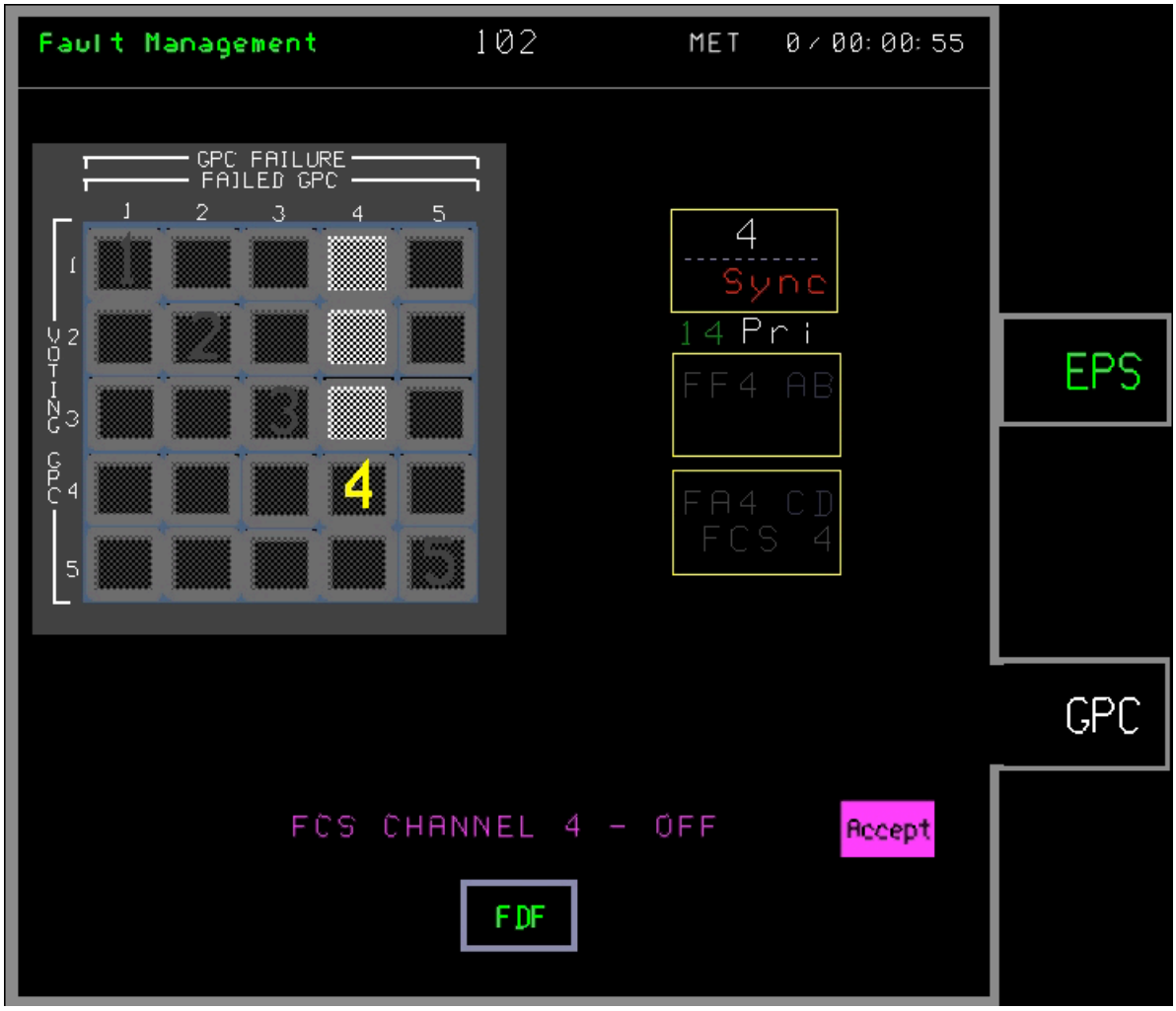


***Figure 2.6**. Fault Management Display for GPC4 fail to synch computer failure. The text section contains the initial procedure from the GPC4 fail to synch section of the FDF to take flight control system (FCS) channel 4 switch to the OFF position. The graphics section consolidates GPC 4 fail to synch information from the overhead GPC matrix and selected string information from the CAU EPS SUM display. Note that the observer has selected the GPC failure by pressing the GPC tab on the right side of the display. The earlier EPS failure has not been resolved (the failure still exists), so the EPS tab remains present and could be selected at any time. That would bring up the Fault Management Page in Figure 2.5.*

2.6 Data Metrics and Modeling

This section describes the collection and analysis of four metrics that collectively form a new approach to evaluating operations concepts, cockpit automation, and cockpit interfaces. The four metrics obtained were performance, situation awareness, workload and eye movements. In addition, this section discusses the modeling approach and results for predicting the performance results.

2.6.1 Fault Management Performance

During each run, the simulator sends various events to the controller computer, such as simulator events (e.g., alarms and malfunction introductions), participant actions (e.g., button presses and switch throws), and some simulator state events (e.g., MET and main engine cutoff). The

controller synchronizes this data with the eye tracking data and a real-time counter. Thus, we are able to measure reaction times of participant responses (e.g., time between an alarm and a button press/switch throw), as well as which procedures were executed and in what order.

For the nonisolatable helium leak, the GPC fail to synch, and the isolatable helium leak (three of the four malfunctions), there is a specific set of manual procedures that must be executed in exact sequential order to resolve the malfunction correctly. Thus, our accuracy calculations define a "correct" resolution of a malfunction using the conservative criteria of all relevant FDF procedures being executed, in the correct order, with no errors of commission. For the fourth malfunction (APC4 subbus failure), no switch throws are required.

An additional measure of performance is fault management response time. Response resolution times were calculated as the time that elapsed from when the malfunction annunciated through the C&W system, to the completion of all applicable procedures as specified in the appropriate section of the FDF.

2.6.2 Situation Awareness

Situation awareness is a measure of a crewmember's understanding of his or her environment. "Good situation awareness" is commonly inferred when a crewmember's actions effect progress toward successful completion of tasks, whereas "poor situation awareness" is inferred when one's actions are not successful. However, it is also possible that one may achieve one's goals (i.e., have good performance) simply through a combination of serendipitous events. To accurately assess situation awareness, we measured crewmembers' understanding of their environment through subjective questions (answered on a rating scale) and objective questions (which have definitive right or wrong answers).

Two examples of the subjective ratings supplied by our participants are shown in Figure 2.7. These were presented to each participant after each off nominal run. Each response was converted to a 1-10 scale for purposes of analysis.

Examples of objective questions answered by participants were:

1. Was the helium leak isolatable (Yes, No)?
2. What aft power controller (APC) sub-bus failed (APC2, APC3, APC4, APC5)?
3. What three subsystems registered impacts of the APC sub bus failure (1. GPC, ECLSS, APU; 2. MPS ULLAGE Pressure, APU, Center Engine He Supply; 3. MPS ULLAGE Pressure, APU, Left Engine He Supply; 4. GPC, APU, Right Engine He Supply)?

The first objective question was asked following the single malfunction run. All three objective questions were asked following the multiple-malfunction run. Participants' ability to answer objective questions was based on memory for events that occurred up to several minutes earlier. While this approach has its drawbacks, we decided not to use the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995), which requires the simulation to halt at intervals so participants can answer situation awareness questions right away, because it disrupts the normal eye movement scan pattern and operational flow of the run. Rather, the questionnaire was given to each participant in paper format immediately following each run. Each participant spent several minutes completing the questionnaire before the next run.

A) Please mark an X somewhere along the following scale to indicate how difficult it was for you to diagnose the malfunctions (i.e., understand the fault messages, locate and process the appropriate display information, etc):

Very Easy — Very Hard

B) How difficult was it to work the malfunctions (i.e., to locate, understand, and execute the appropriate procedures) once you had diagnosed them?

Very Easy — Very Hard

Figure 2.7*. Subjective Situation Awareness questions.*

How useful did you find the "embedded" depiction of the procedure for staying "in synch" with the automation?

Not at all Useful — Very Useful

Please rate the following features of the fault management display:
The schematic representation of the Helium system (for working isolatable and non-isolatable leak procedures).

Not at all Useful — Very Useful

Figure 2.8*. Usability questions.*

Upon completing the four FAMSS data collection runs, participants were given an addition set of questions (called usability questions) to address the overall usefulness of FAMSS and particular features of the Fault Management Display interface. Two examples of usability questions are shown in Figure 2.8. Each response was converted to a 1-10 scale.

2.6.3 Workload

Workload is defined as the mental and physical effort necessary to perform a task. Workload was measured using the Bedford Scale (Roscoe, 1984) and NASA Task Load Index (Hart, 1988). Participants rated their workload on both scales following each run. The Bedford Scale is a 10-point rating scale shown in Table 2.2.

Table 2.2. *Bedford Workload Rating Scale*

	Rating Description
1	Workload insignificant.
2	Workload low.
3	Enough spare capacity for all desirable additional tasks.
4	Insufficient spare capacity for easy attention to additional tasks.
5	Reduced spare ca pacity. Additional tasks cannot be given the desired amount of attention
6	Little spare capacity. Level of effort allows little attention to additional tasks.
7	Very little spare capacity, but maintenance of effort in the primary tasks is not in question
8	Very high workload with almost no spare capacity. Difficulty in maintaining level of effort.
9	Extremely high workload. No spare capacity. Serious doubts as to ability to maintain level of effort.
1 0	Task abandoned. Unable to apply sufficient effort.

NASA TLX is a two-part method for quantifying workload. The first part of TLX is a *rating* of six different workload components (on a scale of 1-20): mental demand, physical demand, temporal demand, performance, effort and frustration. The explanation for each of these components is shown in Figure 2.9. Each participant rated each TLX component following each run.

The second part of TLX is a *weighting*. After all eight data collection runs were complete, participants made pair-wise comparisons by circling which TLX component of each possible pair (such as mental demand versus physical demand) was more important to their experience of workload. With six different components, there were 15 pair-wise comparisons, and each component could be circled anywhere from 0 to 5 times. Once the participants completed these pair-wise comparisons, we tallied the number of times each component was circled (with a possible range of 0 to 5 for each tally); this tally formed the weighting for that component. The overall TLX workload for each run was computed as the sum of each component's rating multiplied by its weighting, and the final TLX workload was scaled from 0.5 to 10.

The immediate benefit of obtaining questionnaire-based metrics such as subjective situation awareness ratings and workload ratings is that they provide a means of assessing a crewmember's opinion of the environment. Further, by correlating subjective metrics (such as workload) with objective metrics (such as response accuracy or eye movement patterns) we can gain insight into how (or weather) specific aspects of task performance translate into subjective impressions of workload and situation awareness.

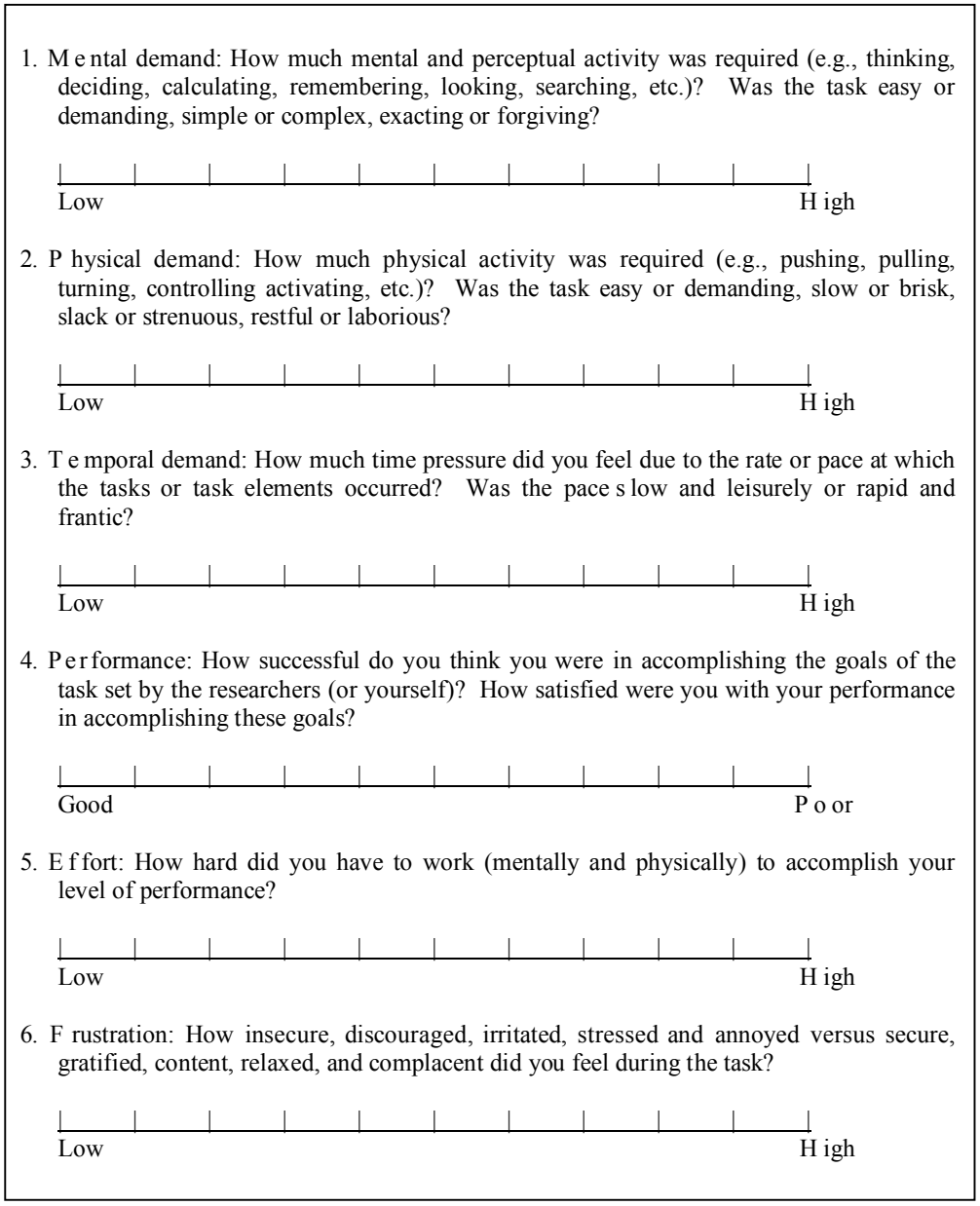


Figure 2.9. *NASA TLX Components.*

2.6.4 Eye Tracking

Objective measures of performance, such as the time it takes an operator to complete a malfunction, are the product of several constituent behaviors such as switch throws, FDF navigation, acquisition of information from display formats, and crosschecks. While the traditional measures can be used to gauge the impact of advanced cockpit interface concepts like FAMSS at a gross level, they provide virtually no information as to the nature and duration of these constituent behaviors. Recording and analyzing eye movements supplements and augments the traditional measures by providing a much more detailed real-time picture of display usage and information acquisition strategies. This more detailed picture helps gauge the usefulness of

individual features or elements of a display design, such as coding and displaying procedural information inside a system schematic. Eye-movement analyses can also be used to quantify the durations of the specific information processing components or physical activities, for example, manual switch throws. These more precise quantities can be used to determine the time savings that could be achieved through automation (though, as we shall see, automation produced benefits over and above what can be predicted merely by subtracting the time needed to perform the automated activity). Eye movements can also be used to evaluate whether a particular design feature has the effect intended by the designer, such as whether the addition of countdown timers and related information to the Fault Management Display succeeded in freeing up participants to engage in more nominal scanning behavior. Thus, eye movement recording and processing is an important part of the ISIS lab tool set in general, and of the FAMSS evaluation in particular. In this section, we describe the technical underpinnings of our eye tracker and our approach to analyzing eye movement data to achieve these design evaluation goals.

2.6.4.1 Calibration Procedures

Because every individual's eyes are unique, the eye tracking system must be calibrated to determine where each person is actually looking. Thus, we began each run by calibrating the eye tracking system using ISCAN's standard procedure:

1. Locate the head in space (called "boresighting") by having the participant orient his head and gaze straight ahead at a known location, and record the head-tracking system's magnetic receiver's output (six degrees of freedom: (x, y, z) and (azimuth, elevation, rotation)) .
2. Determine the relationship between the tracker's video output (corneal reflection and pupil centers: see section 2.2) and the gaze direction by having the participant fixate on five points– straight ahead, ±6.25 degrees of visual angle to the left and right and ±6.25 degrees of visual above and below – while recording the tracker's output.

These procedures produced accurate tracking results only for central display regions adjacent to the calibration points. We found that individual differences and the relative location of the head tracker to the eye affect the accuracy of the calibration and produce artifacts due to fixation location, head position and gaze angle. To minimize these effects, we developed a supplemental calibration procedure, which was used to increase the accuracy of the four most critical displays.

Eye tracker data for each of the four forward liquid crystal display (LCD) displays were collected for a 3x3 grid of fixation crosses and two head positions: leaning forward (lean) and seated upright (no-lean). The calibration procedure consists of two steps for each LCD:

1. Use the measured calibration eye-tracker data for each fixation location (separately for both head positions) to compute the linear transformation which best aligns the measured data with the known calibration locations.
2. Compute mean lean and no-lean head positions.

For the experimental eye tracker data, only the data from the ISCAN planes corresponding to the four forward LCDs and their adjacent ISCAN planes were adjusted by our supplemental calibration procedure. The calibrated location was computed (using the parameters for the

relevant LCD) by applying the linear transformations for each of the head positions, and using the measured head position to linearly interpolate between these values. Typically, this calibration procedure reduced the mismatch between the eye tracker output and the known locations of the calibration crosses by about 2-3 inches for the no-lean data and about 5-6 inches for the lean data.

2.6.4.2 Eye Position Fixation Procedure

Oculomotor research has shown that in searching displays, human eye movements consist of a sequence of gazes at specific locations separated by high velocity, short duration eye movements (saccades) along with head movements. We performed a fixation analysis of the calibrated eye-tracker data to extract the participant's fixation locations and durations. In this study, we are interested only in the fixation locations (normal durations are at least 150 ms), not the details of the rapid saccades (25-75 ms in duration) used to move from one location to another. We used a standard "running average" technique (Duchowski, 2000) to define fixations. The algorithm's input is the eye-tracker horizontal and vertical spatial position data at 60Hz. Each sample is sequentially compared to the current running average. If its spatial location is within a tolerance (horizontal 0.76 inches, vertical 1.5 inches) of the running average, it is added to the running average. This process continues until four consecutive samples fail to meet the spatial criterion (this provides robustness to noise). Then, if the duration is greater than a temporal threshold (120 ms), it is accepted as a fixation, and its start time, duration and average spatial location are saved. Otherwise, the process starts again: the running average is reset to the spatial location of the first point that failed to meet the spatial criterion. This algorithm correctly identified most fixations and successfully rejected saccades. All further eye-tracking analysis used this fixation data.

2.6.4.3 Regions of Interest (ROI) and K-means Clustering Algorithm

In order to understand participants' scanning behaviors, we divided the simulator environment into candidate Regions of Interest (ROI), each of which represented a region where the participants' fixation points typically clustered. After examining the typical clustering patterns in participants' fixation data, the 18 ROI illustrated in Figure 2.10 were defined.

Each fixation point computed by the running average algorithm (described in the previous section) has a corresponding set of (x, y) coordinates, starting time, and duration. To cluster these fixation points into the corresponding ROI, a K-means clustering algorithm (Moody & Darken, 1989) was employed. The K-means algorithm starts with K (18 in our case) sets of initial coordinates, corresponding to cluster centers, and classifies each fixation point as belonging to the cluster whose center is the nearest. Then, the algorithm re-computes new cluster center coordinates by taking the average of all the fixation point coordinates classified into each cluster. The process repeats until no further change in the fixation point grouping is observed.

2.6.5 Human Performance Modeling

2.6.5.1 Rationale

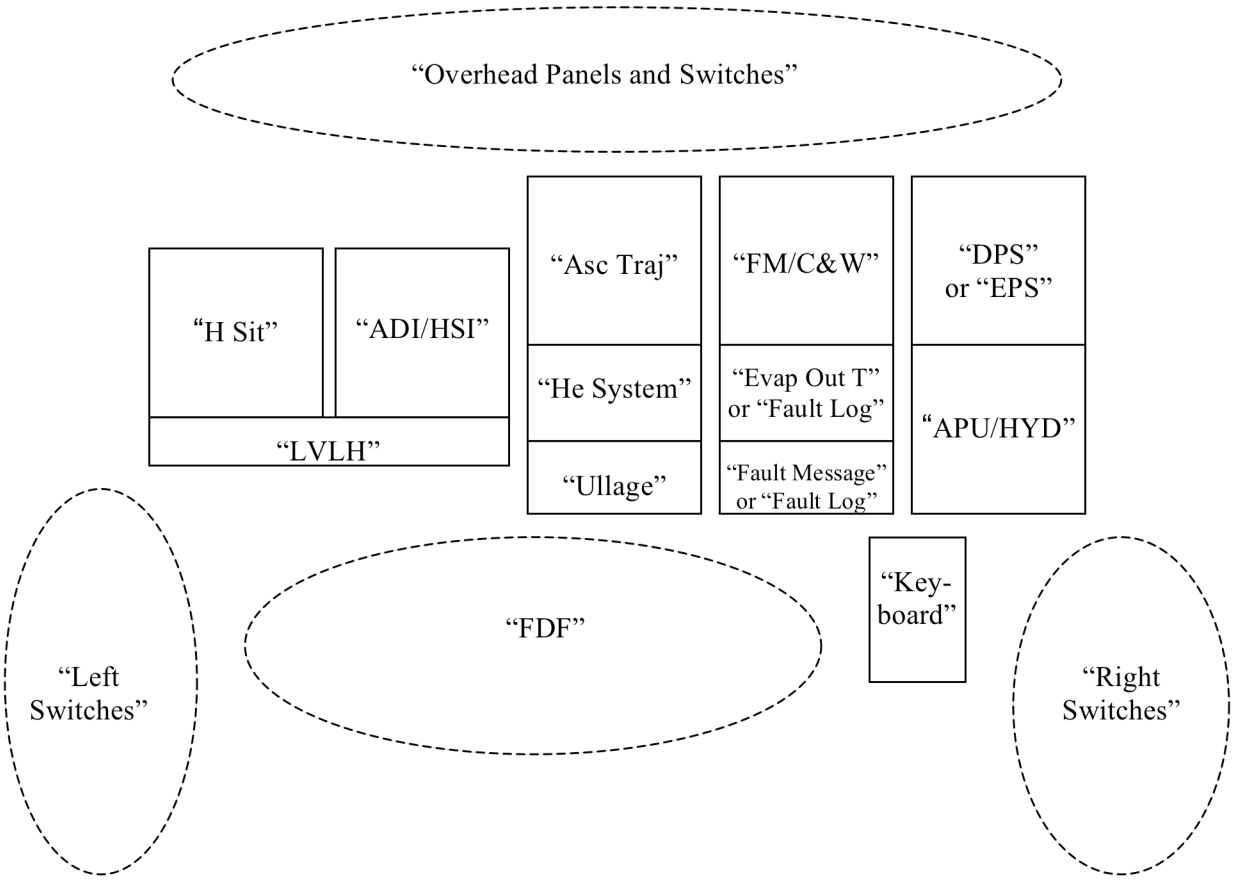


***Figure 2.10**. Eighteen Regions of Interest (ROI)*

In recent years, computational models of human performance have advanced to the point where they are beginning to be able to predict the impact of individual design formats and operational concepts on participants' situation awareness, workload, and operational performance. Accordingly, we are developing a model of human performance in the task environment of a spacecraft cockpit. Our goal is to mature these models to the point where they significantly reduce the requirements for direct human-in-the-loop evaluation of design concepts. Such models will provide the basis for a much more efficient process whereby we evaluate and select among a larger range of rapidly prototyped design concepts via iterative model-based "design-test-redesign-retest" development cycles.

The standard approach to modeling human behavior is to decompose a complex task into a set of primitive operations (templates) to which performance parameters may be assigned. These parameters can be static (e.g., 200 msec for a button press) or dynamic (e.g., a Fitts' Law calculation for mouse movement time based on target distance and target dimensions). These templates represent the building blocks from which entire task sequences can be constructed.

Template reuse is a profitable avenue to explore for a number of reasons. Behavioral templates such as reading text, throwing switches, and checking parameters are parts of many cockpit operations, and templates of these skills do not need to be "built from scratch" for each new cockpit task. Previous empirical validation of reused templates should allow for more accurate

predictions (modeling) of operator performance, such as crewmember latency to work a systems malfunction. Finally, reuse provides additional constraints on models of complex tasks. If the templates predict the behavior well, the task modeler should not change the parameters of the template simply to make it work for a new operation or cockpit user interface.

Previously, the GOMS (Goals, Operators, Methods, and Selection rules; Card, Moran, & Newell, 1983) modeling methodology has been used to decompose complex tasks into a hierarchical set of nested goals and subgoals. A variant of GOMS called CPM-GOMS (John, 1990) creates templates from Cognitive, Perceptual, and Motor operators. CPM-GOMS has been shown to make accurate a priori predictions of human performance in several real-world task domains. An example is Project Ernestine, which predicted the outcome of a test of new computer workstations that saved a telephone company $2 million per year (Gray, John, & Atwood, 1993). Parameterized templates have been created in CPM-GOMS for commonly recurring task-level activities in human-computer interaction, such as mouse moving-and-clicking or typing, which range from a fraction of a second up to several seconds (John & Gray, 1992; Gray & Boehm-Davis, 2000). More germane to our purposes, the GOMS methodology was previously used by Chuah, John, and Pane (1994) to predict times for performing tasks using graphic and textual displays. Their model of comprehending visual information from a single fixation is constructed from an attend-target operator lasting 50 msec, an initialize-eye-movement operator lasting 50 msec, an eye-movement operator lasting 30 msec, a perceive-target operator lasting 290 msec, and a verify-target operator lasting 50 msec. Since each operator begins only upon the completion of the previous operator, the times are additive, giving a total time of 470 msec per individual fixation.

With their further assumption that a fixation can encompass roughly 6 letters in 12-point font, the times for FDF gaze durations can be predicted as follows: reading a key on the keyboard requires one fixation for 470 msec, reading a typical fault message requires two fixations for 940 msec, reading data or a switch label requires three fixations for 1410 msec, and reading a procedure requires eleven fixations for 5170 msec.

The task of interest in this report is real-time fault management. Fault management has five sub-phases: the alerting phase, the fault identification phase, finding and decoding the correct procedure phase, procedure execution phase, and procedure verification phase. Most of the time involved in fault management is a result of visually acquired information, with some time involved in keyboard processing and toggling switches. Therefore, the modeling focused on information acquisition templates with gaze durations and motor response times as parameters. For example, the time for initially finding a malfunction page in the FDF procedures was calculated as gazes at 3.5 (out of 7 possible) systems tabs plus a gaze to read the malfunction heading for a total of (3.5+1) x 470 = 2115 msec. The time for subsequently finding the correct place on the FDF page after looking at the screen was calculated as an orienting gaze to the FDF and a gaze to a finger-keeping place on the page for a total of 2 x 470 = 940 msec. The time for reading information on the screen after looking at the FDF include an extra orienting gaze to the screen.

2.6.5.2 Apex-CPM

Apex-CPM is designed to generate adaptive, intelligent human-like behavior in complex, dynamic environments. Apex incorporates many high-level aspects of cognition including action selection under uncertainty, managing multiple tasks, and scheduling behavioral templates around human resource limitations. Such scheduling is automated in Apex architecture where the agent model makes dynamic decisions about how to allocate its limited resources. In turn, these decisions are influenced by dynamic priorities and policies.

The decision policy in the model is based on simple prioritization schemes. Since handling malfunctions has a higher priority than nominal scanning, a C&W event causes the agent to suspend the nominal scan and attend the malfunction. After the agent performs the necessary steps to resolve the malfunction, it resumes the nominal scanning activities that were interrupted when the alarm sounded.

Our interest was in using the model to make a priori quantitative predictions of the performance impacts that would accrue from working malfunctions in conjunction with the FAMSS interfaces and information sources versus the interfaces and information sources available in the Baseline condition. Making these predictions would then allow us to evaluate the accuracy of the model at its current state of development, and provide guidelines for future improvement. Further, the model functions as an "ideal observer" with no delays or slowdowns to crosscheck information, recover from misreading information, and the many other sources of human variance and error. For example, crosschecks between different sources of information were not captured in the model, whereas such cross checks and verifications are likely to be a feature of human performance. The application of HPM to performance in our particular scenarios enables us to pinpoint actions or activities that are more difficult than the "ideal observer" would predict. In addition, the model predictions act as a sort of "asymptotic" predictor of the minimum improvement we might expect from FAMSS automation with very highly trained operators, such as astronauts.

Effects of Cockpit Condition: Predictions. The phrase-reading and screen-touching templates were used to make predictions for fault management resolution times in the Baseline and FAMSS cockpit conditions for three malfunctions – the isolatable helium malfunction on the single-malfunction run, the GPC fail to synch malfunction on the multiple-malfunction run, and the APC4 malfunction on the multiple-malfunction run. The constituent templates and associated latency parameters for the two malfunctions are itemized in Appendix B. In the case of the APC4 malfunction, there are no overt procedures to complete, so we simply modeled the sequence of information processing and acquisition activities needed to reach the appropriate FDF and read the instruction not to isolate an affected helium supply system. In the case of the other two malfunctions, either three or four procedures have to be executed in correct sequence. The model predictions for the time at which each successive procedure will be completed are shown in Figure 2.11 for the isolatable helium leak and Figure 2.12 for the GPC fail to synch malfunction. In the case of the isolatable helium leak, the model predicts that FAMSS will save approximately 19 sec, or 50% off the predicted 38 sec resolution time for the Baseline Condition. In the case of the GPC fail to synch malfunction, the predicted savings are on the order of 7 sec (33%).

The predicted value for completing the APC4-related activities, approximately 20 sec, is also noteworthy, as that finishing time coincides almost exactly with the C&W event for the GPC fail to synch malfunction. Assuming this prediction is actually an average value, and has variance associated with it, we would expect as many as half of our participants to have not completed APC4-related activities within 20 seconds, and then have to time-share activities on two separate problems. As we shall see shortly, time-sharing was indeed common, and had very interesting consequences for the efficiency with which the GPC malfunction was handled.

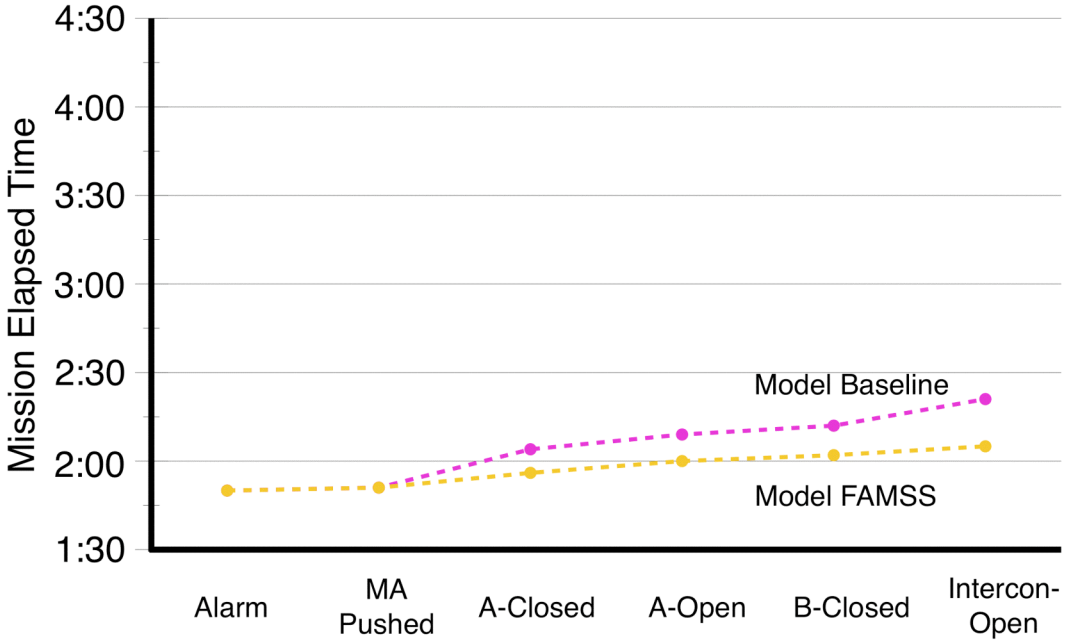


Figure 2.11. *APEX-GOMS Model predictions of procedure completion times for the isolatable helium leak as a function of Cockpit Condition. Baseline Condition is in pink; FAMSS condition in yellow.*

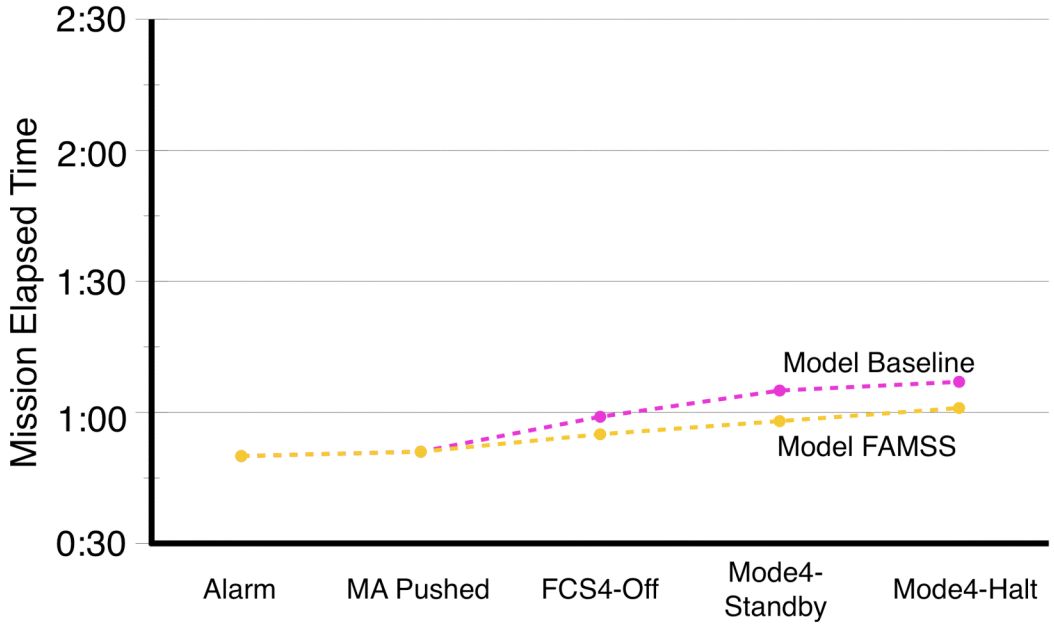


Figure 2.12. *APEX-GOMS Model predictions of procedure completion times for the GPC4 fail to synch malfunction as a function of Cockpit Condition. Baseline Condition is in pink; FAMSS condition in yellow.*

3 Results

3.1 Scenario Level

We first
present our results at the scenario level (single malfunction versus multiple malfunction) to provide a "big-picture" overview of malfunction handling errors as a function of scenario complexity and cockpit condition. Then, in the following sections, we proceed with more fine-grained analyses at the individual malfunction level (e.g., GPC fail to synch) to obtain further insight into the sources of fault-management difficulty and FAMSS impacts.

3.1.1 Errors in Malfunction Management Performance

As specified in the AESP FDF, correct fault management for three of the four malfunctions requires specific procedural actions (recorded as button pushes or switch throws) in a specific order, with no extraneous responses (errors of commission). Using these very stringent criteria for fault management accuracy, the percentage of off-nominal runs resolved incorrectly is shown in Figure 3.1. For the single-malfunction (isolatable helium leak) scenario, participants were unable to resolve the malfunction correctly on 43% of the Baseline Condition runs; that percentage dropped to zero on the FAMSS runs. For the multi-malfunction scenario, participants were unable to correctly resolve all malfunctions on 72% of the Baseline Condition runs, compared to 22% of the FAMSS runs.

A three-way split-plot ANOVA with Condition Order (Baseline on Day 1, FAMSS on Day 2 versus FAMSS on Day 1, Baseline on Day 2) as the between-subjects effect and Cockpit Condition (Baseline or FAMSS), and off-nominal scenario complexity (Single versus Multiple

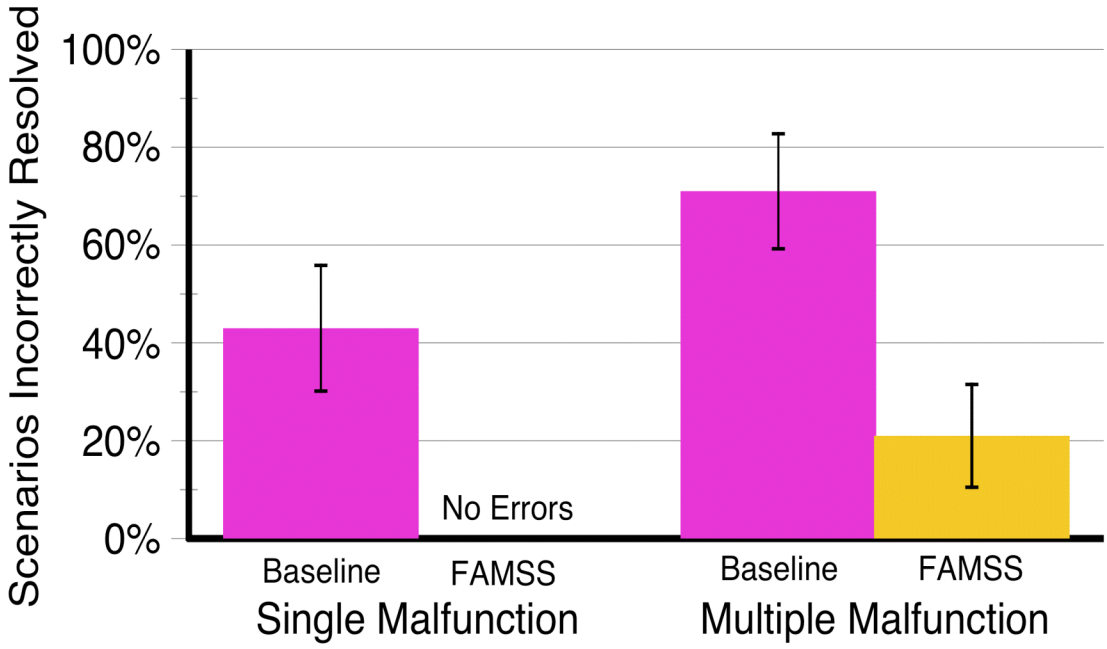


Figure 3.1. *Percentage of scenarios resolved incorrectly by condition (Baseline versus FAMSS) and scenario complexity (single malfunction run versus multiple malfunction run).*

malfunction) as within-subject effects revealed a significant main effect of Cockpit Condition, $F(1,12)=21.1$, $p < 0.01$, with fewer scenarios resolved incorrectly in the FAMSS condition (3/28 scenarios or 11%) than in the Baseline condition (16/28 scenarios or 57%). There was also a significant effect of scenario complexity, $F(1,12)= 14.7$, $p < 0.01$, with fewer single-malfunction scenarios resolved incorrectly (6/28 scenarios or 21%) than multiple-malfunction scenarios (13/28 scenarios or 46%). There was no significant effect of cockpit order ($p = .8$), which implies that any practice effects on accuracy were small compared to the effects of cockpit and malfunction count. There were no significant interactions.

The error numbers presented above require perfect performance of all procedures on all malfunctions. To more precisely define the effects of cockpit condition, we examined errors at the level of the individual malfunctions (more analyses results regarding the effects of cockpit condition will be presented in the next section, 3.2.) Because the single-malfunction scenario contained only the single isolatable helium leak, the errors on the malfunction itself is the same as the errors on the scenario as a whole. Performance for each malfunction in each cockpit condition is shown in Figure 3.2.

Summing across participants, 19 of 42 malfunctions (45%) were not completed correctly in the Baseline condition, compared to 3 malfunctions (7%) in the FAMSS condition. Broken out by individual malfunction, the GPC fail-to-synch failure was not resolved correctly on 43% of the Baseline runs compared to 14% of the FAMSS runs, the nonisolatable helium leak was not resolved correctly on 50% of the Baseline runs and 7% of the FAMSS runs, and finally, as noted in the previous section, the isolatable helium leak was resolved incorrectly on 43% of the

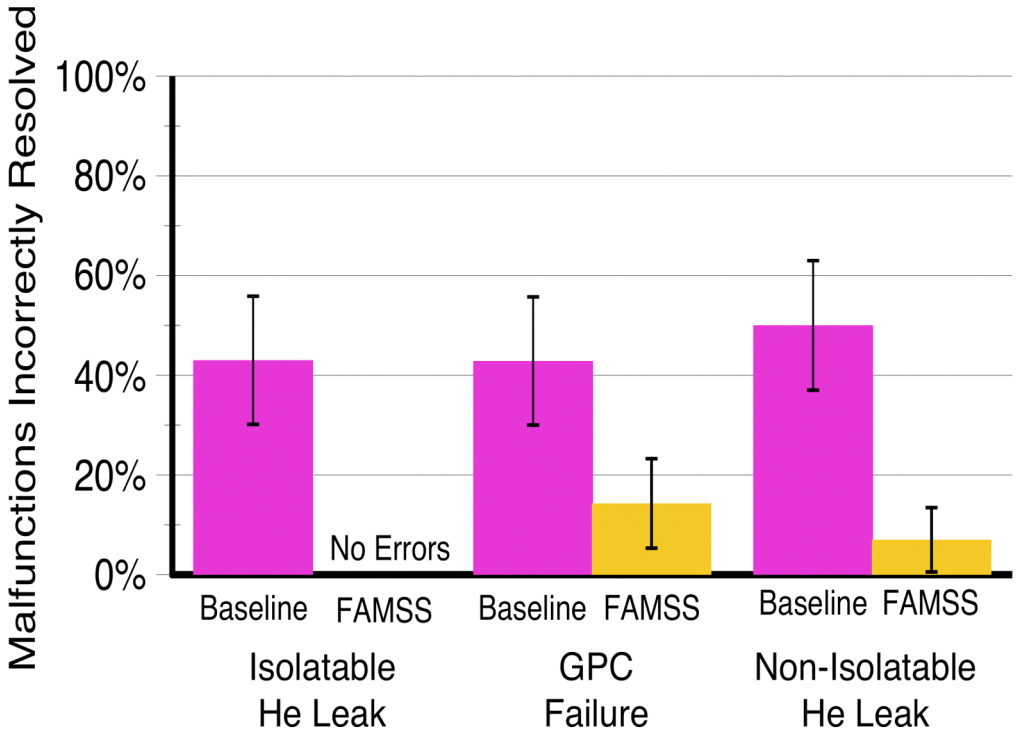


Figure 3.2. *Percentage of malfunctions resolved incorrectly by condition (Baseline versus FAMSS).*

Baseline runs compared to none of the FAMSS runs.

A split-plot ANOVA with Condition Order (Baseline first versus FAMSS first) as the between-subject effect and Cockpit Condition (Baseline versus FAMSS) and Malfunction (GPC fail to synch, isolatable helium leak, and nonisolatable helium leak) as within-subject effects revealed a highly significant effect of Cockpit Condition (Baseline less accurate than FAMSS), $F(1,12) = 18.73$, $p < .01$. No other main effects or interactions were significant.

3.1.2 Situation Awareness and Usability

Malfunction-related situation awareness was measured by asking participants to rate their "ability to diagnose the malfunction," and "ability to resolve the malfunction" on a 1-10 scale. Average ratings for each Scenario and Cockpit Condition are shown in Figure 3.3. Separate ANOVAs were conducted on each rating with Condition Order (Baseline on Day 1 versus FAMSS on Day 1) as the between-subjects effect and Cockpit Condition (Baseline or FAMSS), and off-nominal scenario complexity (Single versus Multiple malfunction) as within-subject effects. As might be expected, participants found the malfunctions significantly easier to diagnose in the single-malfunction scenario than in the multiple-malfunction scenario, $F(1,12) = 4.9$, $p < 0.05$. Participants also found it to be easier to diagnose malfunctions in the FAMSS cockpit than in the Baseline cockpit, $F(1,12) = 29.8$, $p < 0.01$. There was no effect of Order and no significant interactions.

Similar results were obtained on the question asking participants to rate their ability to resolve the malfunction once it was diagnosed. Participants found it significantly easier to resolve malfunctions in the single-malfunction scenario, $F(1,12) = 9.4$, $p < 0.01$, and significantly easier

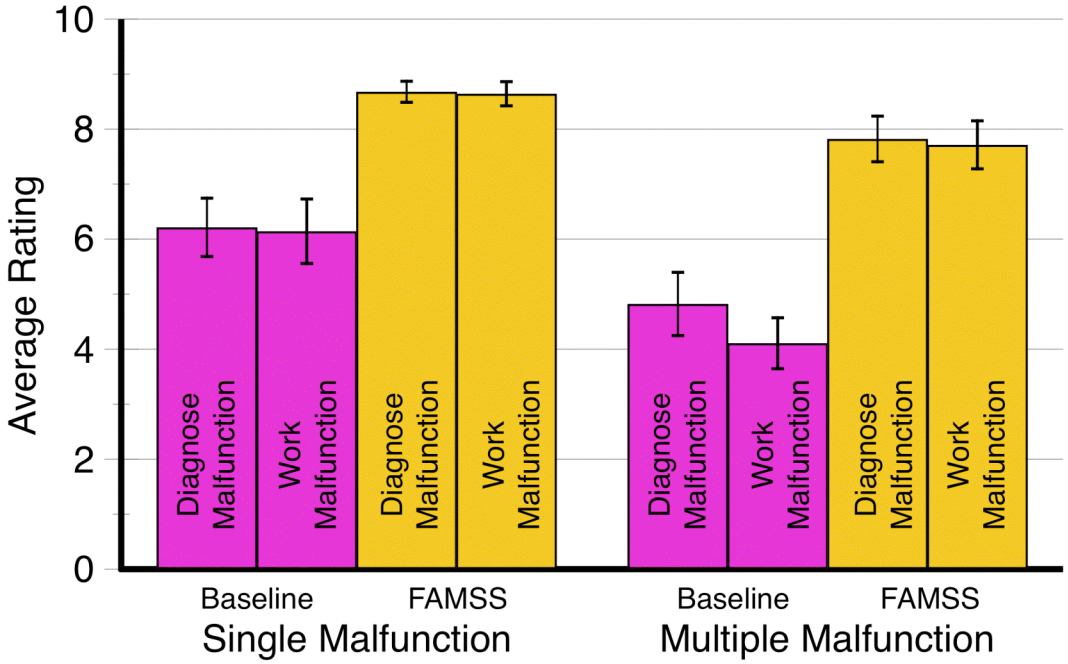


Figure 3.3. *Rated ability (scale of 1 to 10) to diagnose and resolve malfunctions by condition (Baseline = pink bars; FAMSS = yellow bars) and Scenario Complexity (single versus multiple malfunction runs).*

to resolve malfunctions in the FAMSS cockpit, $F(1,12) = 31.4$, $p < 0.01$. Again, there was no main effect of Order and no significant interactions.

Participants clearly found it easier to diagnose and resolve the malfunctions in the more highly automated (FAMSS) condition; however, increased automation often comes at the expense of reduced operator understanding of the operation (Endsley & Kiris, 1995; Billings, 1997). For example, in the FAMSS condition, participants could have resolved the malfunctions by accepting the procedural recommendations on the Fault Management Display without taking the time to read and understand those recommendations (see Section 3.3.1 for eye movement results addressing this issue). To test that participants actually understood the operations they were performing, several objective questions were asked about the malfunctions (such as, "Was the He leak isolatable?"; see Section 2.6.2). Responses to three questions were analyzed for the Multiple-malfunction condition and one for the Single-malfunction condition. Performance on these questions was identical across the Baseline and FAMSS conditions (73% correct). These results indicate that participants at least felt that they understood the actions performed by the automation about as well as when they performed those actions themselves.

At the end of the entire experiment, participants were asked to rate 13 specific features of the FAMSS Fault Management Display (e.g., coding of procedures within helium supply system schematics, color coding and countdown indicators) on a scale of 1 (not at all useful) to 10 (extremely useful). All features were rated very highly. The average rating was 8.9 and no feature was rated lower than 7.5. Every feature was rated a 10 by at least two participants (one participant rated all 13 features a 10). Only two features were rated as having medium to low usefulness by more than one participant. In particular, the engine schematic that appeared prior to engine shutdown during the nonisolatable helium leak was rated below 5 by two of the 14 participants and below 7 by five participants. The next most poorly rated feature was "text for automation," rated below 5 by one participant and below 7 by another. No other feature was rated below 7 by more than one participant.

3.1.3 Workload

Two measures of workload were used in this experiment: the Bedford Scale and NASA Task Load Index (TLX). Mean ratings from each of these methods are shown in Figure 3.4. The within-subject effects of Cockpit Condition (Baseline versus FAMSS) and Scenario Complexity (Single versus Multiple malfunction runs), and the between-subjects effect of Order (Baseline on Day 1 versus FAMSS on Day 1), were assessed using separate ANOVAs for each workload measure.

The Bedford Scale ANOVA showed significant main effects of Cockpit Condition, $F(1,12) = 16.3$, $p < 0.01$ and Scenario, $F(1,12) = 19.7$, $p < 0.01$ but no interactions. As with situation awareness, there was no significant effect of Order or interactions. Planned comparisons were conducted to assess the effect of Cockpit condition on Bedford Scale ratings in the single and multiple-malfunction scenarios separately. On the multiple-malfunction run, workload was rated 37% lower (5.6 versus 3.5) in the FAMSS condition, $t(13) = 3.3$, $p < 0.01$. On the single-malfunction run, workload was rated 27% lower in the FAMSS condition (3.5 versus 2.6; $t[13] = 2.6$, $p < 0.05$).

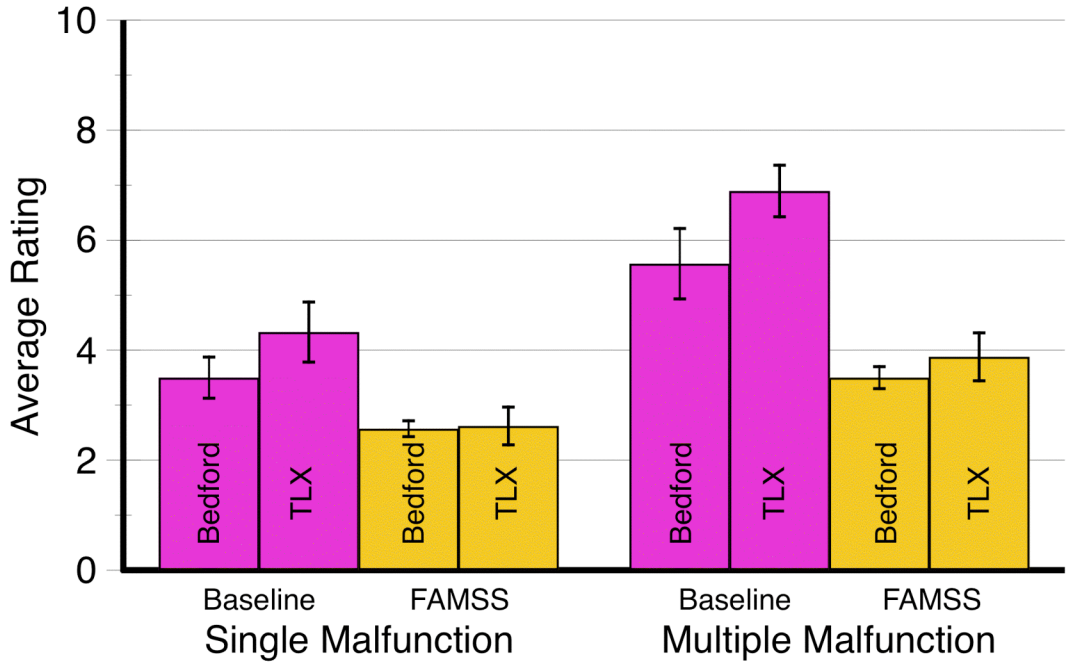


Figure 3.4*. Average Workload Ratings (Bedford and TLX) by Cockpit Condition (Pink Bars = Baseline, Yellow Bars = FAMSS) and Scenario Complexity (Single Malfunction Runs on left, Multiple Malfunction Runs on right).*

NASA TLX results were very similar to Bedford results. Both the main effect of Cockpit Condition (Baseline versus FAMSS) and Scenario (Single versus Multiple Malfunction) were significant, *F's*(1,12) = 62.9 and 15.0, respectively, both *p*'s < 0.01). Unlike the Bedford analysis, there was a marginally significant interaction between these variables, $F(1,12) = 3.4$, $0.05 < p < 0.10$, reflecting the fact that cockpit condition had a larger effect on the multiple-malfunction run than on the single-malfunction run. Planned comparisons revealed that workload ratings were significantly lower for the FAMSS condition on both the multiple-malfunction run (6.9 versus 3.9; $t[13] = 4.9$, $p < 0.01$), and the single-malfunction run (4.3 versus 2.6; $t[13] = 3.9$, $p < 0.01$). As with previous analyses, there was no effect of Order and no significant interactions.

Workload ratings were also analyzed for the nominal runs. TLX workload was slightly higher for Baseline than FAMSS (2.3 versus 1.5; $F[1,12] = 15.3$, $p < 0.01$). Similar trends were found for Bedford workload, although the results were only marginally significant, $F(1,12) = 4.7$, $0.05 < p < 0.10$. These results are notable because the actual manual requirements are identical in the two conditions (throwing an attitude indicator switch to the "LVLH" position 9 sec into launch). The information acquisition requirements (ascent checklist item monitoring and nominal scanning) were identical in the Baseline and FAMSS conditions, and the display format layout was virtually identical, with only a slight shift in the position of the CAU Fault SUM display.

There are three possible sources for this intriguing result. It may be a general context effect of operating (in one case) in an environment where a great deal of support was provided for

working malfunctions, and in another case (Baseline) where most of the work was the responsibility of the operator. The second possibility is that the result is a local "anchor" effect, such that when workload was rated high on the proceeding off-nominal trial, this "dragged" the rating on the subsequent (nominal) trial higher too. The third possibility is that the result reflects a genuine difference in participants' information acquisition and processing activities on nominal trials that depended on the context.

3.2 Baseline versus FAMSS Comparison at Individual Malfunction Level

We now turn to an extensive series of analyses at the individual malfunction level (isolatable helium leak, GPC fail to synch, APC4 subbus failure, and nonisolatable helium leak) to further understand how the participants' performances and scan patterns were affected by cockpit condition during different malfunction management performance periods. We will start with the isolatable helium leak, and then the GPC fail to synch. These two malfunctions required the participants to throw appropriate switches in a specific order, and that enabled us to explicitly define the malfunction resolution time (i.e., from the Master Alarm to the last switch throw). For these malfunctions, the errors in performance, the malfunction resolution time, display fixation pattern changes, inter-procedure intervals, and finally comparison with the model-predicted performance will be examined. Next, we present results for the APC4 subbus failure case, where no overt procedure (e.g., switch throws) was required so the only available measures were eye movements and videotapes. Lastly, we present results for the nonisolatable helium leak. The two procedures associated with this malfunction were under the temporal control of external events, such as the Center Engine helium supply system reached a particular tank pressure, and thus, the inter-procedure intervals did not show many differences between Baseline and FAMSS conditions. Therefore, for this malfunction, only accuracy analyses and fixation patterns on cockpit displays are included.

3.2.1 Isolatable Helium Leak

3.2.1.1 Errors in Fault Management Performance

The isolatable helium leak was the only malfunction on the run. The malfunction required pressing the C&W master alarm button (or FAMSS software equivalent) followed by four manual procedures defined in the AESP FDF. An analysis of the percentage of participants who failed to perform each successive procedure correctly provides direct insight into what aspects of the fault management process were the most challenging, and where FAMSS produced the greatest benefits.

The cumulative errors during the isolatable helium malfunction management performance reveal a progressive pattern of performance degradation across procedures in the Baseline condition (Figure 3.5). Ignoring the silencing of the master alarm, one participant failed to complete any of the four procedures correctly. This is not to say that this participant made no attempt to work the fault; to the contrary, he actually closed both ISOL B and ISOL A, shutting the engine down. An additional participant took the first step (ISOL A CL) correctly, but left ISOL A in the center (GPC) position when attempting to open it back up. When he subsequently closed ISOL B, both valves were closed, again shutting the engine down.

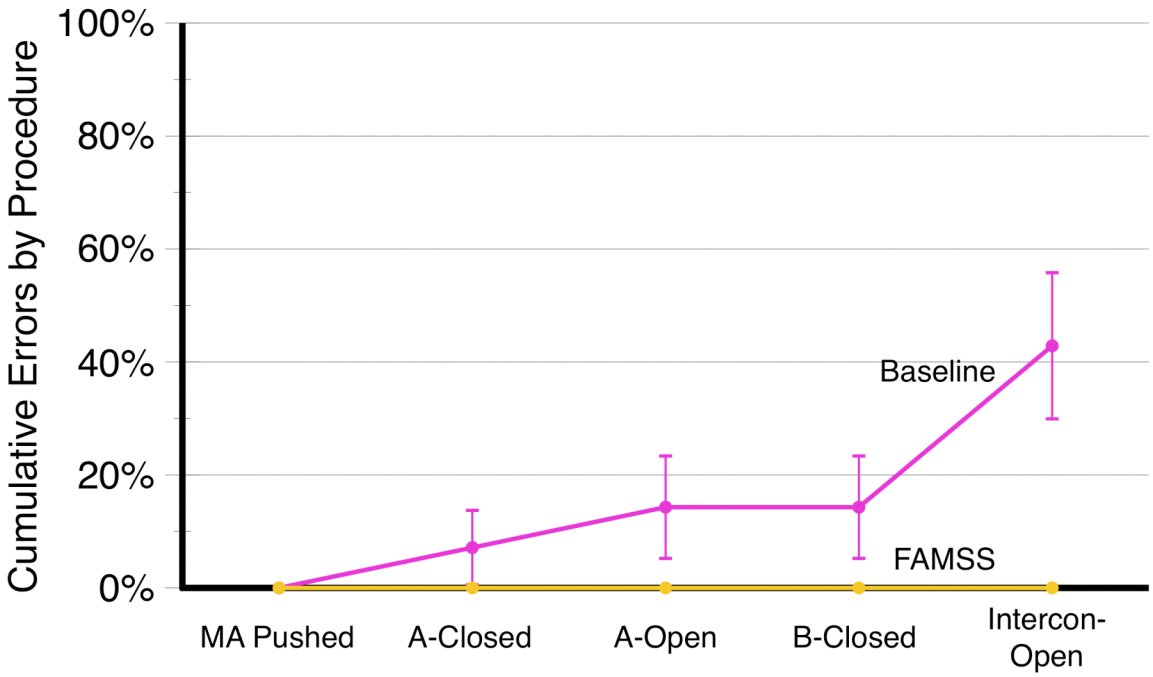


Figure 3.5. *Cumulative error rate across successive procedures for the isolatable helium leak (single malfunction scenario). MA Pushed = Master Alarm press; A-Closed = Right Engine helium supply system ISOL A switch to closed position; A-Open = Right Engine helium supply system ISOL A switch to open position; B-Closed = Right Engine helium supply system ISOL B switch to close position; Intercon-Open = Right Engine helium supply system interconnect switch to IN-OPEN position.*

The remaining 12 participants completed the three isolation procedures correctly. However, an additional four participants (29%) failed to perform the final procedure (opening the interconnect), despite extensive training (some as recent as the morning of the day the data was collected) on correct FDF navigation for the isolatable helium case. These errors of omission are quite interesting, as we know from our earlier description of the FDF that the interconnect procedure was perceptually segregated from the isolation procedures, with several intervening navigation steps and logical conditionals. Inspection of the videotape recordings made it very clear that these participants thought they had completed all required procedures when they finished isolating the leak. We have more to say about this issue in the next section.

3.2.1.2 Malfunction Resolution Time

For the analysis of the malfunction resolution time, we included only those participants who resolved the isolatable helium leak malfunction correctly, and computed malfunction resolution time as the time from the appropriate C&W event to the completion of all required FDF procedures. Resolution times for the isolatable helium leak are shown in the right side of Figure 3.6. The isolatable helium leak was resolved much faster in the FAMSS condition (Mean = 46 sec), and with much less variability, than in the Baseline Condition (Mean = 128 sec). A two factor ANOVA including Cockpit Condition and Cockpit Presentation Order effects revealed a significant effect of Cockpit Condition, $F(1,6) = 8.8$, $p < 0.05$. There was no main effect of condition order and no interaction.

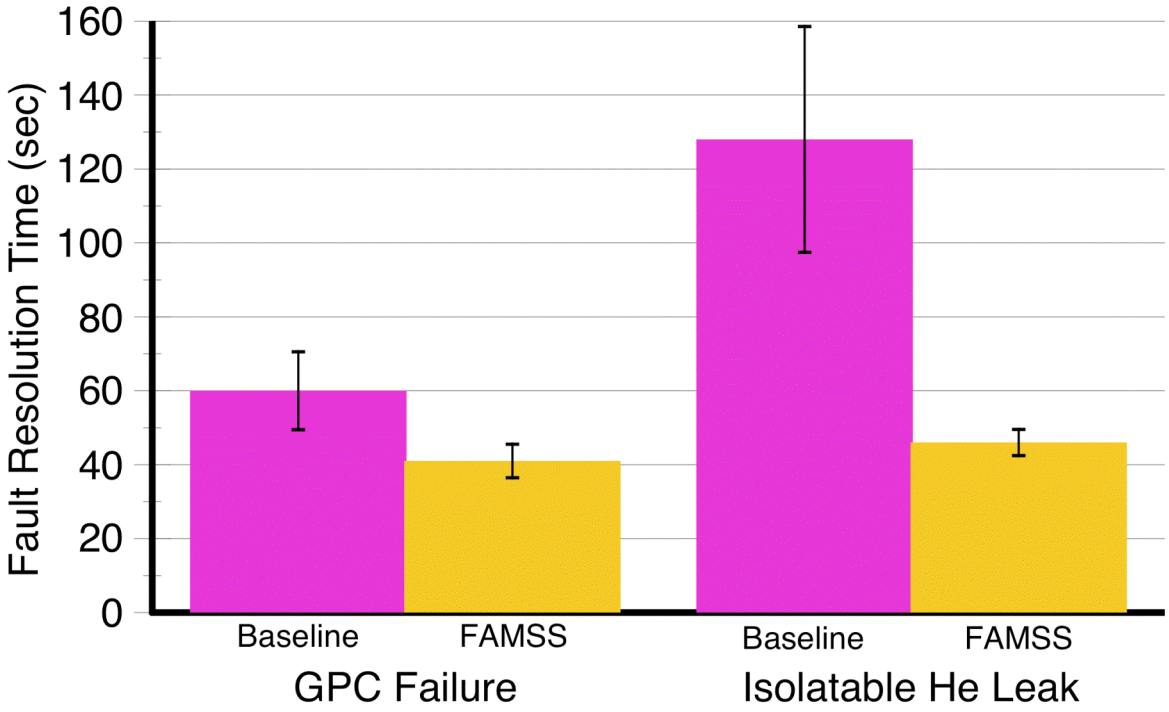


Figure 3.6*. Malfunction resolution times for the GPC fail to synch malfunction (left side) and the isolatable helium leak malfunction (right side).*

3.2.1.3 On-Task versus Off-Task Time by Eye Fixations

The previous section showed that the eight participants who correctly completed the fault management procedures for the isolatable helium leak completed them much more quickly in the FAMSS condition than in the Baseline condition. In this section, these participants' fixation data are further examined to make a direct empirical connection between this speedup and the FAMSS Fault Management Display.

For this analysis, the fixation data from the helium leak alarm time to 4 seconds after the last switch throw of the procedure were examined. The 4 seconds were added at the end because the isolatable helium leak page of the Fault Management Display in the FAMSS condition stays on for approximately 4 seconds after the last switch throw. Although the Baseline condition does not have the Fault Management Display, the same 4 seconds were added after the last switch throw in the Baseline Condition so that the comparison would be fair. Two of the eight participants had relatively large percentages of missing eye-movement data (> 40%) during the isolatable helium leak fault management period of their Baseline run, and were not included in this analysis.

The first step in this analysis was to compute the total time spent processing information directly related to the isolatable helium leak malfunction, or *on-task* time. In order to compute the on-task time, the total fixation durations on Regions of Interest (ROI) related to the isolatable helium leak (i.e., "He System," "Fault Management Display/C&W," "Fault Message," "Fault

Log," "FDF [Baseline condition]," "Keyboard," and "Right Switches" as shown in Figure 2.10) were computed. Then, an ANOVA with Cockpit Condition and Cockpit Presentation Order as main effects was performed on these on-task times. The results showed that the on-task time during the isolatable helium leak malfunction management period was significantly shorter in the FAMSS Condition (Mean = 30 sec) than in the Baseline Condition (Mean = 71 sec), $F(1,4) = 53.34$, $p < 0.01$. Shorter on-task time in the FAMSS cockpit is consistent with the faster malfunction resolution in the FAMSS condition than in the Baseline condition.

To better understand what contributed to the difference in on-task times in the two cockpit conditions, we further examined the components of the on-task time. First, the total fixation durations only on the "Fault Management Display" for the FAMSS condition were compared with the total fixation durations on the "Fault Message" and "FDF" for the Baseline condition, since, in the FAMSS condition, the Fault Management Display consolidates the information provided on the Fault Message part of the Fault Summary Display and on the paper FDF in the Baseline condition. The analogous ANOVA was applied, and the results showed that the sum of all fixation durations on the Fault Management Display was significantly shorter (Mean = 20 sec) than the total fixation durations on the "Fault Message" and "FDF" combined (Mean = 35 sec), $F(1,4) = 32.26$, $p < 0.01$. The result shows that similar information was processed much faster when represented and displayed on the Fault Management Display in the FAMSS Condition than when represented and displayed across different locations and formats of the Baseline Condition.

Second, the total fixation durations on the "Right Switches" and the "He System" in the on-task ROI were examined. Remember that, although the Fault Management Display in the FAMSS cockpit allows the operator to resolve the isolatable helium leak malfunction without looking at these ROI, participants were encouraged to crosscheck the Fault Management Display with these ROI. The analogous ANOVA results indicated that the total fixation durations on these ROI in the FAMSS cockpit were significantly shorter (Mean = 8) than those in the Baseline cockpit (Mean = 36 sec) ($F(1,4) = 25.81$, $p < 0.01$). The result makes sense because in the Baseline condition, participants would have needed to fixate on these regions much longer in order to look for the switch, throw the switch, and confirm it on the MPS SUM, while in the FAMSS condition, they would just crosscheck these regions and would not need to fixate on the switch panel and the MPS SUM for as long. Thus, this effect is also considered to have contributed to faster completion of the isolatable helium leak procedures.

Interestingly, in the Baseline condition, five of the six participants fixated on the "FDF" ROI for an average of 48 sec after they completed the isolatable helium leak procedures correctly, presumably to verify that they completed all procedures. On the other hand, the Fault Management Display in the FAMSS cockpit indicates "SYSTEMS NOMINAL" after all procedures are completed, eliminating the need for continuing verifications that distract from nominal scanning activities.

In addition to the on-task time, the *off-task time* was also computed as the total fixation durations on ROIs not directly related to the malfunction management task (i.e., for the isolatable helium leak procedures, "H Sit," "ADI/HSI," "Asc Traj," "Ullage," "Evap Out T," "DPS," "EPS," and "APU/HYD"). Two out of the eight participants who correctly completed the isolatable helium

leak procedures (one of the two was among those whose data were not included in the fixation analyses due to the large percentage of the missing data during this period) showed extremely long off-task time during the isolatable helium leak malfunction management procedures in the Baseline condition (70 seconds or more). Naturally, they were also the slowest to complete the procedures. Such long off-task times were not observed in the FAMSS cockpit. The inter-procedure interval data for these participants indicated that these participants had extremely long intervals between the second to the last switch throw and the last switch throw. The results suggest that they may have thought that the procedures were complete after they finished with the isolation procedures, and gone back to nominal scanning. We will revisit this issue in the next section.

3.2.1.4 Inter-Procedure Intervals

Next, we further examined participants' performance by calculating the inter-procedure intervals (i.e., the time elapsed from completion of procedure X to completion of Procedure X+1). For the isolatable helium malfunction, the mean MET at which each individual procedure was completed is shown in Figure 3.7, and the mean inter-procedure intervals are shown in Figure 3.8. Like the GPC problem, the initial inter-procedure interval (time to complete the first procedure in the sequence of leak isolation steps) showed a large effect of cockpit condition. In sharp contrast to the GPC malfunction, however, the largest FAMSS benefit was on the final procedure (taking the interconnect valve to "IN-Open").

These observations were supported by statistical analyses. An ANOVA on inter-response interval with Procedure and Cockpit Condition as independent variables revealed a main effect of Cockpit Condition (Baseline slower than FAMSS), $F(1,7) = 8.6$, $p < .05$, and a main effect of Procedure, $F(3,21) = 4.5$, $p < .05$. Most critically, the interaction of Procedure and Cockpit Condition was significant, $F(3,21) = 3.6$, $p < .05$, reflecting the fact that FAMSS reduced inter-procedure intervals to a greater extent for the initial (Open A) and last (Interconnect OP) procedures than for the middle procedures.

3.2.1.5 Comparison with Model Predictions

Figure 3.9 repeats the model predictions From Figure 2.11 along with actual results for the isolatable helium leak malfunction. The comparisons reveal an interesting pattern for the isolatable helium malfunction. Looking at Condition only, we see that the model greatly underestimated the actual time it took our participants to work the malfunction. The underestimate is noticeable for the first procedure, but what really stands out was just how much the model underpredicted the latency to work the final interconnect procedure. The pattern is similar to other results that show FAMSS had its greatest impact on these procedures.

In fact, the pattern is quite understandable. Clearly, progressing from the isolatable set of procedures to the interconnect command (Step 10 in Figure 1.4) was far more difficult than predicted. Participants either failed to navigate the FDF at all (in which case they missed the procedure) or they failed to navigate through to Step 10 immediately, but came back to the procedure later on, or they navigated correctly but at half the rate predicted by this "ideal observer." That is clearly why FAMSS had its biggest impact here, as FAMSS automated the process of FDF navigation.

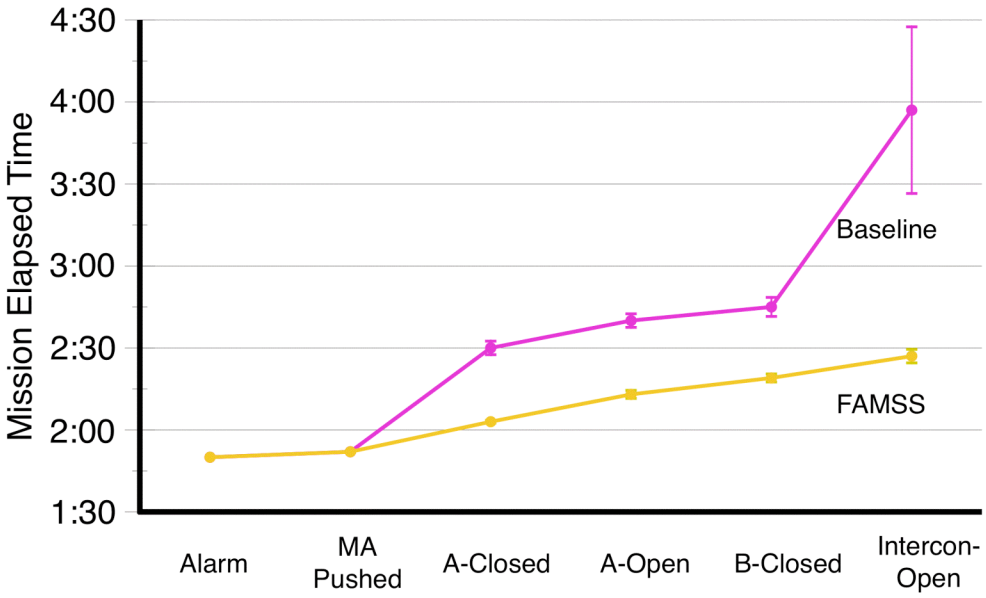


Figure 3.7. *Mean completion time for isolatable helium leak (single malfunction scenario) procedures in Baseline and FAMSS conditions. MA Pushed = Master Alarm press; A-Closed = Right Engine helium supply system ISOL A switch to closed position; A-Open = Right Engine helium supply system ISOL A switch to open position; B-Closed = Right Engine helium supply system ISOL B switch to close position; Intercon-Open = Right Engine helium supply system interconnect switch to IN-OPEN position.*

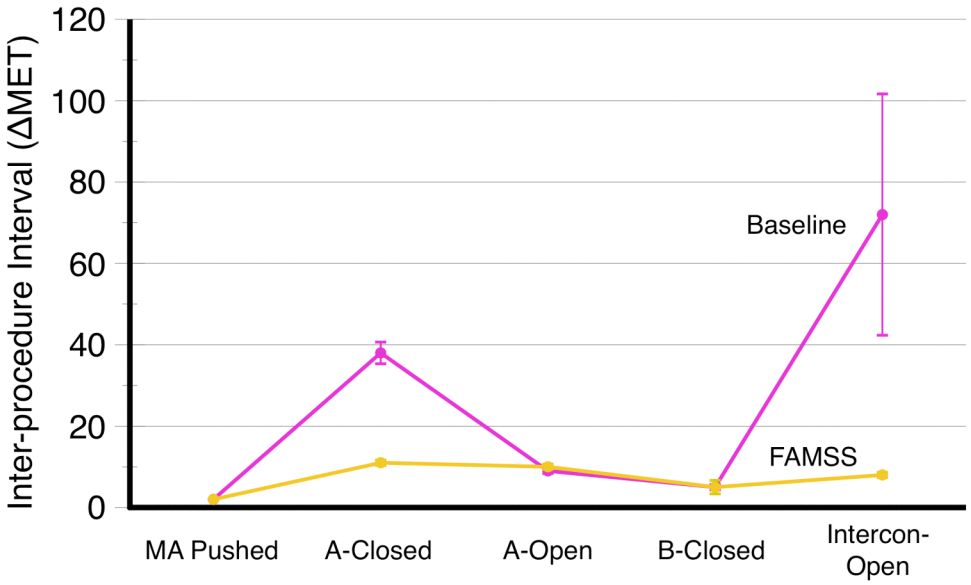


Figure 3.8. *Mean inter-procedure intervals for isolatable helium leak in Baseline and FAMSS conditions. MA Pushed = Master Alarm press; A-Closed = Right Engine helium supply system ISOL A switch to closed position; A-Open = Right Engine helium supply system ISOL A switch to open position; B-Closed = Right Engine helium supply system ISOL B switch to close position; Intercon-Open = Right Engine helium supply system interconnect switch to IN-OPEN position.*

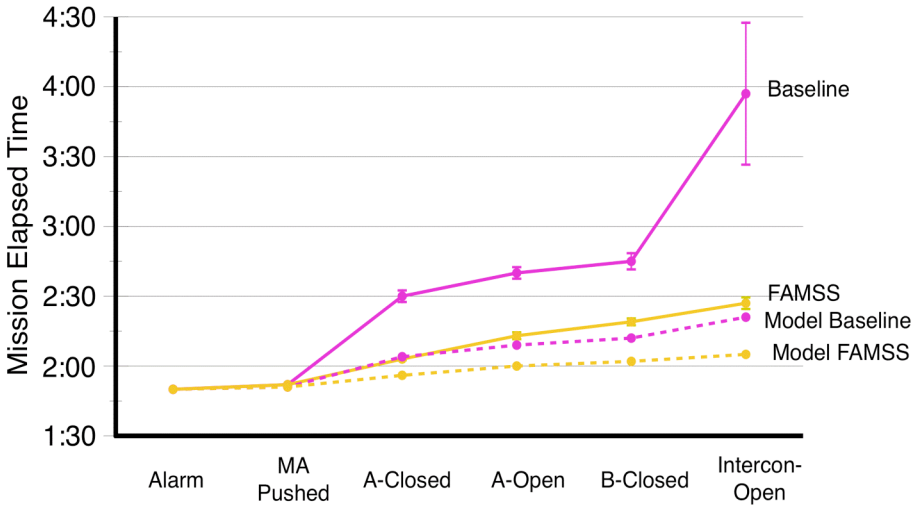


Figure 3.9. *Actual versus predicted procedure completion times for the isolatable helium leak. Baseline Conditions are in pink; FAMSS conditions are in yellow. Dotted lines are model predictions; solid lines are actual results.*

3.2.2 GPC Fail to Synch

3.2.2.1 Errors in Fault Management Performance

Turning to the GPC fail to synch problem (Figure 3.10), although all participants acknowledged the problem, they either proceeded to work all successive procedures correctly, or none of them correctly, in both cockpit conditions. However, the source of the problem was different in the two cockpit conditions. Two participants failed to resolve the problem correctly in the FAMSS Condition compared to seven participants in the Baseline Condition. In the Baseline Condition, two members of this group did not work the problem at all, out of an erroneous conclusion that it was a consequence of the earlier APC4 subbus failure. One member of this group understood the FDF instructions, completed the first two procedures correctly, and actually physically attempted the final procedure (take the GPC mode switch from "Standby" to "Halt"). However, his physical attempt failed to move the switch past "Standby", and he failed to crosscheck the final switch position against the talkback indicator, leaving the computer in "STBY" mode. The three remaining members of this group also completed all three actions prescribed in the FDF correctly, but made errors of commission with the first (FCS4-OFF) procedure. In addition to taking FCS4 to "Off," these participants toggled some or all of the remaining three FCS switches to the "Off" position.

In the FAMSS condition, by contrast, the two participants failed to push the "GPC" tab on the Fault Management Display, and so never replaced the APC4 failure page with the GPC Fail-to-Synch page. Thus, theirs were errors of omission, rather than commission.

3.2.2.2 Malfunction Resolution Time

As in the resolution time analysis for the isolatable helium leak, we included only those participants who resolved the malfunction correctly, and computed malfunction resolution time

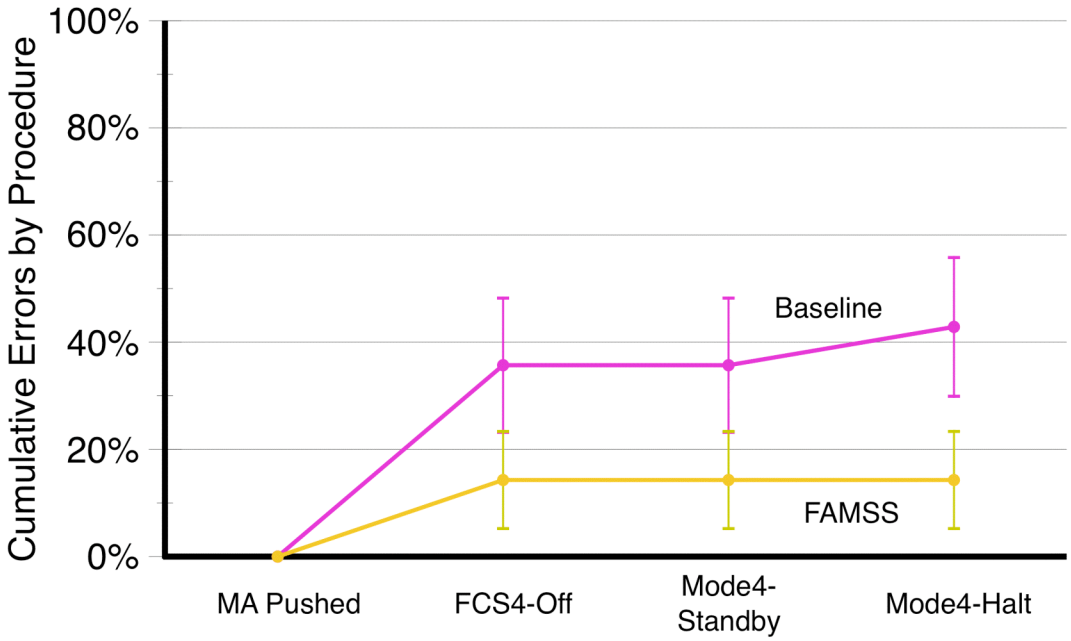


Figure 3.10. *Cumulative error rate across successive procedures for the GPC fail to synch malfunction. FCS4 = flight control switch; MODE4 = general purpose computer 4 mode control switch on Panel O2.*

as the time from the appropriate C&W event to the completion of all required FDF procedures. Resolution times for the GPC malfunction are shown in Figure 3.6. For the GPC fail to synch problem, average malfunction resolution time was 60 sec in the Baseline Condition and 41 sec in the FAMSS Condition. A two factor ANOVA with Cockpit Condition and Cockpit Presentation Order effects revealed no significant main effects or interactions. We will have more to say about these results in the upcoming analyses of inter-procedure intervals.

3.2.2.3 On-Task versus Off-Task Time by Eye Fixations

The fixation analysis time period for this malfunction was from the time when the participant appeared to have started to work on the GPC fail to synch on videotape recording (Baseline) or the time the GPC tab on the Fault Management Display was pushed for the first time (FAMSS), to 4 seconds after the completion of the last switch throw. The 4 seconds were added for the same reasons as in the isolatable helium leak analysis case.

The on-task times (ROI associated with the GPC fail to synch malfunctions: “Fault Management Display/C&W,” “Fault Message,” “Fault Log,” “DPS,” “APU/HYD,” “FDF” (for Baseline condition only), “Keyboard” and “Overhead Panels and Switches”) were computed for the five participants who correctly completed the GPC fail to synch malfunction management procedures in both cockpits, and also looked up the FDF during the GPC fail to synch malfunction management procedures in the Baseline cockpit (two participants performed the entire GPC fail to synch procedures from memory, and were not included in the analysis even though they correctly completed the procedures). The ANOVA results with Cockpit Condition and Cockpit Presentation Order effects showed that the on-task time during the GPC fail to synch malfunction

management procedures was significantly shorter in the FAMSS cockpit (Mean = 19 sec) than in the Baseline cockpit (Mean = 33 sec), $F(1,3) = 24.28$, $p < 0.05$.

To examine what contributed to the on-task time difference in the two cockpit conditions, first, the total fixation durations on the "Fault Message" and "FDF" for the Baseline cockpit or the "Fault Management Display" for the FAMSS cockpit were computed and compared. However, the ANOVA did not find any significant effect. Then, the total fixation durations on the "Overhead Panels and Switches" were computed and subjected to the same ANOVA. The results showed that the total fixation durations on the "Overhead Panels and Switches" were marginally shorter in the FAMSS cockpit (Mean = 4 sec) than in the Baseline cockpit (Mean = 9 sec, $F(1,3) = 6.00$, $p = 0.092$. Thus, the shorter fixation durations on the "Overhead Panels and Switches" ROI, rather than those on the "FDF," "Fault Message," or the "Fault Management Display," were the likely contributors to the shorter total on-task time in the FAMSS condition.

Likewise, the off-task times (ROI: "H Sit," "ADI/HSI," "Asc Traj," "He System," "Ullage," Evap Out T," APU/HYD," and "EPS") were computed and the same ANOVA was applied. However, no significant effects were found.

3.2.2.4 Inter-Procedure Intervals

For the GPC malfunction, the average MET at which each GPC fail to synch procedure was recorded is shown in Figure 3.11, and the inter-procedure-intervals (mean elapsed time between completion of procedure X and procedure X+1) are shown in Figure 3.12. By far, the longest inter-procedure-interval (40 sec) was associated with the initial action (FCS4 to "Off") in the Baseline Condition. This is also the procedure showing the largest benefit for the FAMSS condition. Once the procedures commenced, they progressed at roughly similar rates in the two cockpit conditions.

Inspection of Figure 3.12 suggests the presence of a statistical interaction between Cockpit Condition and Procedure. However, an ANOVA that included these two variables revealed only a significant effect of Procedure, $F(2,12) = 25.9$, $p < .01$, with the FCS4 procedure taking much longer to initiate than taking the appropriate GPC mode switch on Panel O2 to STBY and then HALT. This is hardly surprising, as the last two procedures involved the same mode control switch. But the lack of an interaction, and indeed, the lack of a main effect of Cockpit Condition on overall malfunction resolution time are both interesting in their own right. The lack of statistical significance on an effect apparently as large as this is an indication that considerable variability across participants is lurking under the average interval.

That was clearly the case here. The standard error for the inter-procedure interval in the Baseline Condition (10.7 sec) is considerably larger than for any other condition. Indeed, the standard error in the figure substantially underestimates the true variability in the inter-procedure interval time, because the figure only includes participants who performed *all* GPC fail to synch procedures accurately. Recall from the previous "Errors in Malfunction Management Performance" section (Section 3.2.2.1) that one participant was excluded from these analyses only because he failed in his attempt to take the GPC Mode control switch from STBY to HALT. He completed the critical first procedure correctly, and if we are just interested in the inter-procedure times for that procedure, there is no good rationale to exclude him. Three additional

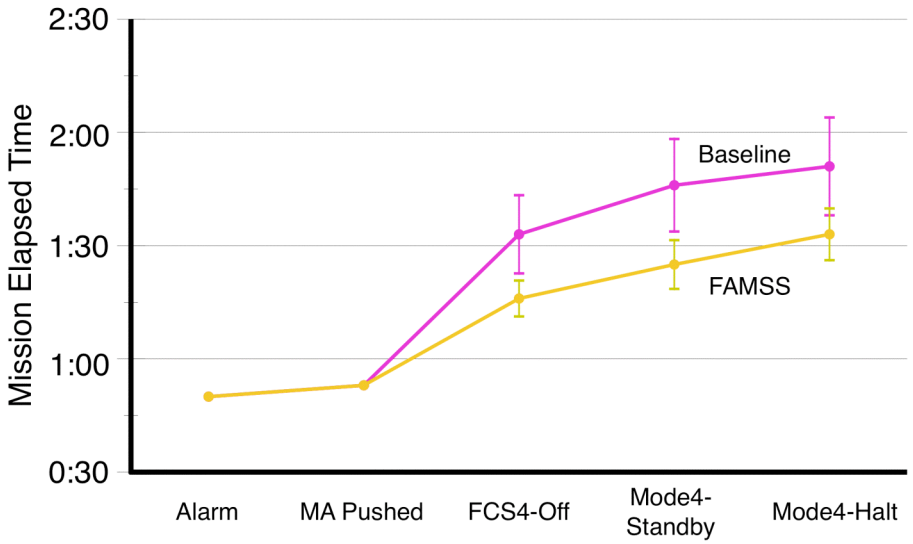


Figure 3.11. *MET for individual procedures in Baseline and FAMSS conditions for the GPC fail to synch malfunction. MA Pushed = Master Alarm Press, FCS4-Off = Flight Control System Switch #4 to "Off", Mode4-Standby = GPC Mode Control Switch on panel O2 to "Standby"; Mode4-Halt = GPC Mode Control Switch on panel O2 to "Halt".*

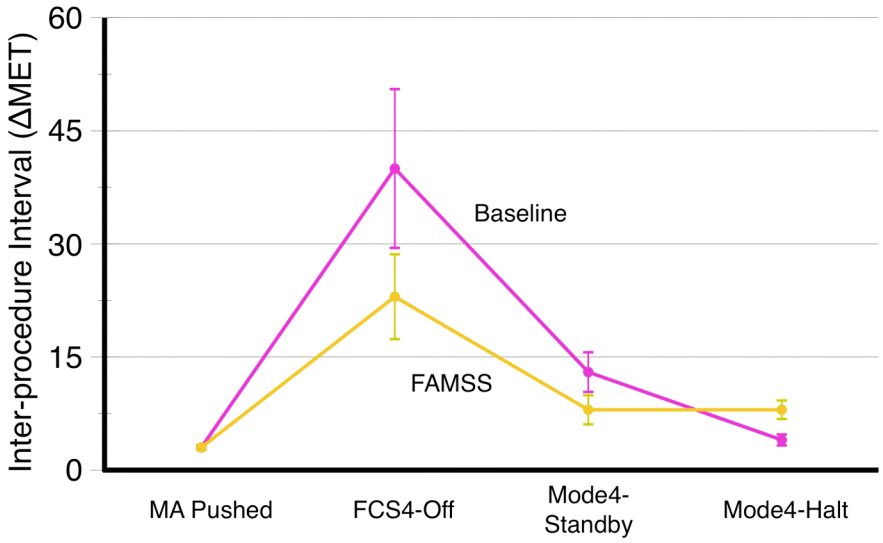


Figure 3.12. *Mean Inter-procedure Interval for GPC 4 fail to synch malfunction by cockpit condition. MA Pushed = Master Alarm Press, FCS4-Off = Flight Control System Switch #4 to "Off", Mode4-Standby = GPC Mode Control Switch on panel O2 to "Standby"; Mode4-Halt = GPC Mode Control Switch on panel O2 to "Halt".*

participants made errors of commission on the FCS switches but did take FCS4 to the off position, and proceeded to complete all remaining procedures correctly. If we include them all in the inter-procedure interval calculation, the mean interval increases to 72 sec (versus 22 sec in the FAMSS condition), and the standard error doubles to 20 sec (as compared to 4 sec for FAMSS; $F(9,9) = 4.7$, $p < .05$).

What was the source of this variability, and why was it so much larger in Baseline than in FAMSS? Recall that at the time the GPC4 malfunction was inserted, participants had had only 20 seconds to process the complex APC4 subbus malfunction, with its multiple off-nominal cockpit signatures and C&W fault messages. Our model-based predictions for the time to process APC4 provided a strong hint that some participants would not have completed APC4-related processing when the GPC failed. One straightforward hypothesis, then, is that the high variance in inter-procedure interval for the first GPC procedure was due to individual differences in how much they engaged in APC4-related processing after the GPC failure was inserted.

Analyses of eye movement data in conjunction with participants' verbal callouts from videotapes of the data collection runs provided the means to test this hypothesis. For each participant who completed the first GPC procedure correctly, we quantified the amount of time spent fixating on APC4-related regions of the cockpit after the GPC fail to synch Master Alarm occurred (i.e., the GPC fail to synch start time used in the on-task versus off-task time analysis minus 20 sec MET). We then regressed the GPC inter-procedural interval times against the APC4 time. The scatter-plot and linear regression line are shown in Figure 3.13.

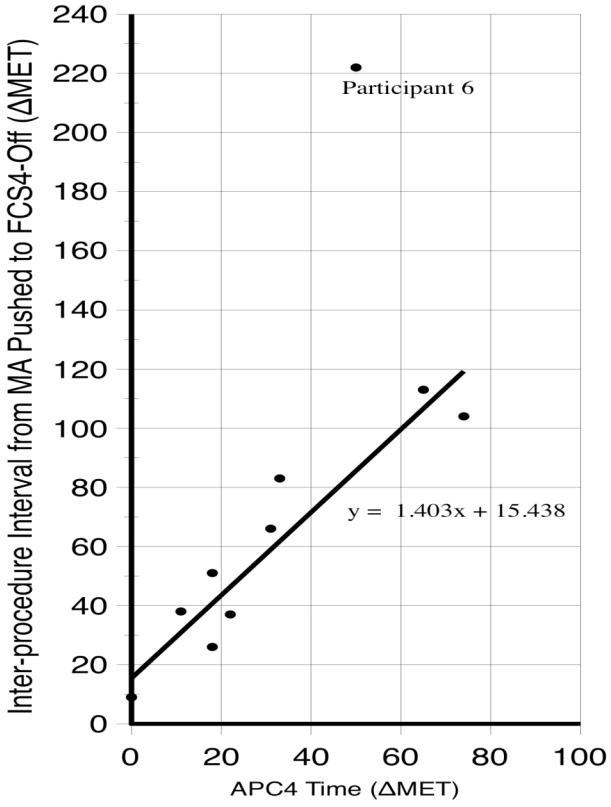


Figure 3.13*. Latency to perform the first AESP FDF procedure (FCS4-Off) for GPC4 fail to synch malfunction as a function of summed fixation durations on APC4-related regions of interest after the GPC4 malfunction occurred. Participant 6 was a special case who neglected the GPC4 problem until much later in flight.*

The results are clear. Virtually all of the variance in the time it took to progress from pressing the GPC4 fail to synch alarm to toggling the FCS4 switch to "Off" is accounted for by the duration of APC4-related processing. The correlation was 0.94, with a very strong linear dependence between the variables. The slope of the regression line was 1.4, $t(8) = 8.1$, $p < .01$.

3.2.1.5 Comparison with Model Predictions

Figure 3.14 repeats the model predictions from Figure 2.12 along with the actual results for inter-procedure times for the GPC fail to synch malfunction.

As in the isolatable helium leak case, the model is worse at predicting the initial procedure behavior than the rest of the procedure. The average difference between model and data is 31 seconds for the first switch throw and 3 seconds for the remaining two switch throws. For both GPC fail to synch and the isolatable helium leak malfunctions, inspection of videotapes revealed that this underestimation was associated with "page flipping"; several participants took much longer to locate the correct section and subsection in the FDF than predicted by the model. Again, FAMSS had the biggest impact here because FAMSS automated that activity.

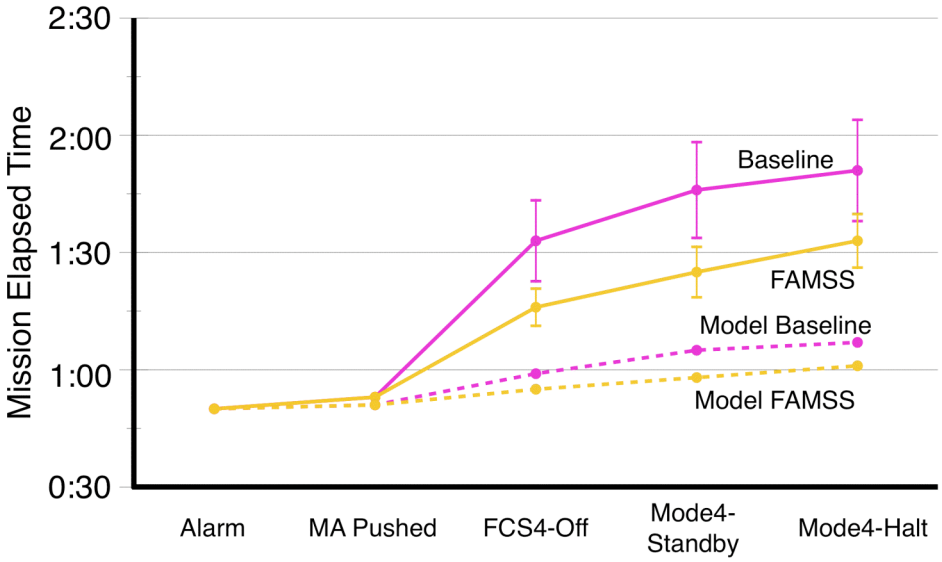


Figure 3.14. *Predicted and actual inter-procedure times in Baseline and FAMSS conditions for the GPC4 fail to synch malfunction. MA Pushed = Master Alarm Press, FCS4-Off = Flight Control System Switch #4 to "Off", Mode4-Standby = GPC Mode Control Switch on panel O2 to "Standby"; Mode4-Halt = GPC Mode Control Switch on panel O2 to "Halt".*

3.2.3 APC Subbus Failure

3.2.3.1 On-Task versus Off-Task Time by Eye Fixations

The on-task time from the APC4 alarm time (30 seconds MET) to the time the participant started to work on the GPC fail to synch malfunction (the same time used for the GPC fail to synch on-task/off-task time analysis) were analyzed. As mentioned before, the eye-movement data were

the only available objective measures for this APC4 subbus failure malfunction management period, as these malfunction management procedures do not require any physical switch throw activities of the operator. Thus, only the results of the eye-movement data analysis are presented for this malfunction period.

The on-task ROI for this malfunction were "He System," "Ullage," "Fault Management Display/C&W," "Fault Message," "Fault Log," "EPS," "APU/HYD," "FDF" (for Baseline condition only) and "Keyboard." As in the isolatable helium leak analysis, an ANOVA with Cockpit Condition and Cockpit Presentation Order as main effects was applied to the on-task time of ten participants who at least started working on the GPC fail to synch malfunction so that their GPC fail to synch starting time could be identified. However, no significance was found in their on-task times during the APC4 period.

Then, the analogous ANOVA was also applied to the ten participants' off-task times. The off-task ROI for APC4 malfunction were "H Sit," "ADI/HSI," "Asc Traj" and "DPS." The ANOVA found that the off-task time during the above-mentioned APC4 malfunction management period was marginally shorter in the FAMSS cockpit (Mean = 8 sec) than in the Baseline cockpit (Mean = 13 sec), $F(1,8) = 4.93$, $p = 0.057$. Finding shorter off-task time in the FAMSS cockpit may sound counterintuitive and require explanation. In the Baseline cockpit, the participants usually had to read the Fault Message and then go to the proper page in the FDF. This process took time, and also the participants often inserted monitoring the off-task ROIs in between examinations of Fault Message and the FDF. On the other hand, the Fault Management Display in the FAMSS condition provided the malfunction management warning ("Do not isolate MPS He C") adjacent to the Master Alarm push button, minimizing the chance for the participant to look away from any on-task related ROI between the Master Alarm and the Fault Management Display looks. Thus, the off-task times in the Baseline cockpit tended to be longer than those in the FAMSS cockpit.

3.2.4 Nonisolatable Helium Leak

3.2.4.1 Errors in Fault Management Performance

In the Baseline condition, accuracy was at 50% for the nonisolatable helium leak on both procedures, which simply indicates that seven participants failed to perform either deferred procedure correctly. One participant spent several minutes after the first alarm of the trial (the APC4 malfunction) flipping through the AESP in search of a procedure. He never did find the procedure he was looking for, and ultimately did not attempt any procedure in response to either of the two active malfunctions. The remaining six participants failed to navigate to the nonisolatable section of the MPS He P checklist. Importantly, all six initially started to work the problem as an *isolatable* helium leak. Four of these failed to catch the error before closing both ISOL A and ISOL B and causing the Center Engine to shut down. What is particularly remarkable about this behavior is that in all but one of these cases, the participants had responded to the earlier APC4 problem with overt verbal annunciations to the effect that, if a helium problem occurred later in flight, they should not attempt to isolate it!

For the FAMSS condition, the one participant who failed to resolve the nonisolatable problem correctly misunderstood the contingency conditions, and did not give FAMSS permission to

execute the engine shutdown procedure. Instead, he watched the countdown indicator "time out" and then assumed that the engine shutdown was completed automatically.

3.2.4.2 Malfunction Resolution Time

The timing of the two deferred procedures for the nonisolatable helium leak was highly constrained by external events (i.e., tank pressure reaching a certain level, vehicle reaching a certain speed). Not surprisingly, therefore, resolution times for the nonisolatable helium leak hardly differed between the Baseline and FAMSS conditions (7 min 33 sec and 7 min 26 sec, respectively).

3.2.4.3 On-Task versus Off-Task Time by Eye Fixations

The fixation analyses time period for this malfunction was from the start time of the nonisolatable helium leak malfunction management procedures (obtained from videotape recording) to 4 seconds after the completion of the last switch throw. Again, the 4 seconds were added for the same reasons as in the isolatable helium leak analysis and the GPC fail to synch failure cases.

The on-task time (ROI associated with the nonisolatable helium leak malfunctions: "ADI/HSI," "He System," "Fault Management Display/C&W," "Fault Message," "Fault Log," "FDF [for Baseline condition only]," "Keyboard," and "Right Switches") and the off-task time (ROI: "H Sit," "Asc Traj," "Ullage," " Evap Out T," "DPS," "EPS," and "APU/HYD") during this period were computed for the six participants who correctly completed the nonisolatable helium leak procedures in the both cockpits, and also started the nonisolatable helium leak procedures within a reasonable time range in both cockpits (one participant's data had to be omitted from this analysis because he did not start the procedures until very late in his FAMSS trial, making it difficult to directly compare his FAMSS data with his Baseline data). An ANOVA found no significant effect on both the on-task and off-task time.

However, the ANOVA results of the total fixation durations on the "Fault Message" and "FDF" ROI for the Baseline condition or the "Fault Management Display" for the FAMSS condition indicated that fixation durations on these ROIs were significantly longer in the FAMSS condition (Mean = 84 sec) than in the Baseline condition (Mean = 64 sec), $F(1,4) = 8.02$, $p = 0.05$ (see Figure 3.15). The total fixation durations on the "He System" and the "Right Switches" ROI related to the cross-checking the procedures and helium system state, on the other hand, showed opposite trends, being significantly shorter in the FAMSS Condition (Mean = 45 sec) than in the Baseline Condition (Mean = 70 sec), $F(1,4) = 28.14$, $p < 0.01$. The opposite trends between the total fixation durations on the two groups of on-task ROIs likely made the comparison of on-task time between cockpit conditions not significant. As in the isolatable helium case, we expected to observe fewer total fixation durations on the cross-checking ROIs ("He System" and "Right Switches") in the FAMSS cockpit, where the participants did not have to look for the switches and physically throw them. By contrast, the fact that total duration of fixations on the Fault Management Display ROI in the FAMSS condition was longer than in the Baseline condition was opposite to the intent of the designers. As mentioned above, there are two switch throw steps in the nonisolatable helium leak procedures that need to be performed after certain vehicle system parameters reach a specified value (e.g., the Center Engine helium supply system tank pressure reaching 1140 psi), and the Fault Management Display was designed to assist the

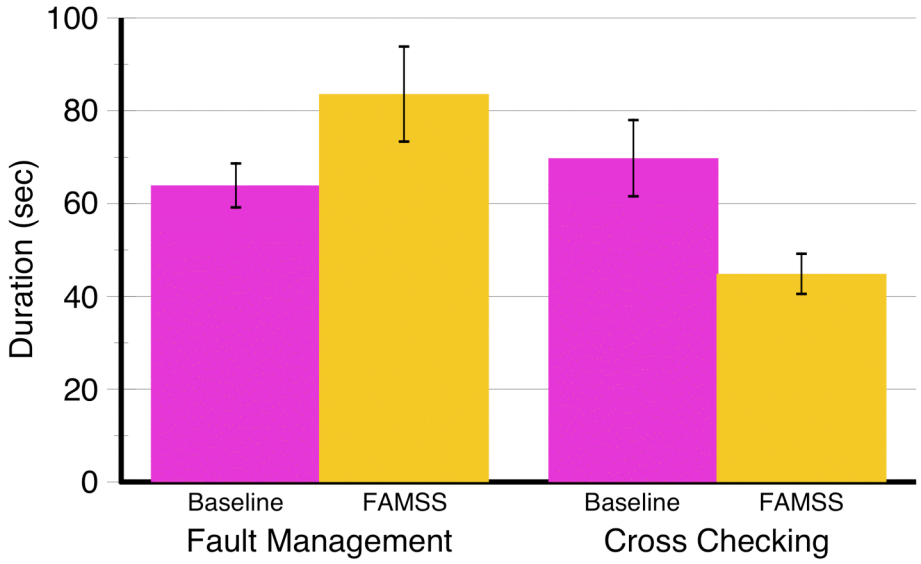


Figure 3.15. *Averages and standard errors of total Fault Management ROI (Fault Message and FDF for Baseline; Fault Management Display for FAMSS) fixation durations (left) and total Cross Checking ROI (He System and Right Switches) fixation durations (right) during nonisolatable helium leak management period.*

participants to determine how much time they had before the deferred procedures needed to be executed (via the countdown indicator). Our goal was for participants to use this information to better manage their information acquisition strategies during the deferred period, and devote a greater fraction of their time to nominal scanning. However, the countdown indicator and other features of the Fault Management Display may have had the opposite effect, causing cognitive tunneling and longer total fixation durations on the Fault Management Display than on equivalent ROIs in the Baseline condition.[1]

3.3 Evaluation of Fault Management Display Features

In this study, the most important display for evaluation purposes was the new Fault Management (FM) Display in the FAMSS condition. This section focuses on three more detailed FM Display usage questions:

1) How much time did the participants spend using the Fault Management Display while working malfunctions?
2) Did participants utilize FM schematic information?
3) What were participants' transition patterns between various FM Display features and other regions in the cockpit that contained fault-relevant information? What do these transition patterns tell us about how participants coordinated and integrated their information acquisition strategies?

To address these questions, eye-movement data for each of the four malfunctions were analyzed separately; each included data from the time the master alarm sounded until the time the

[1] Interestingly, however, in both the single and multiple-malfunction runs, more participants completed their earlier T-MECO guidance convergence check (Table 2.1) in the FAMSS cockpit (57%) than in the Baseline Cockpit (43%)

malfunction was resolved and the FAMSS display returned to its nominal state[2]. Only data from participants who successfully completed the malfunction in their FAMSS run were included (not being restricted by participants' performance in the Baseline condition enables us to include more participants' data in the analyses in this section). The FM display was sub-divided into three additional ROIs, a top schematic portion, a bottom text portion, and a right- side tab region (see Figure 1.5 for an example of the nonisolatable helium leak FM display). Each fixation on the FM display was assigned to one of these three ROIs.

3.3.1 Fault Management Display Usage

The first issue we were interested in was whether participants chose to utilize the FM Display or they continued to solve malfunctions in the traditional way by using the paper FDF, the relevant displays, and the hard switches. Although we assumed that participants would prefer the FM Display to these "legacy" regions, the eye-movement data analysis is the only way to directly verify this assumption and to quantify how much time the participants actually spent using the FM Display. In fact, the eye movement (EM) data show that in solving each malfunction, instead of looking at the paper FDF (no fixations), the participants relied heavily on the FM Display (as shown by the large amount of time spent fixating it). The mean (across participants) total times (and standard errors) spent fixating the FM Display for each of the four malfunctions were: isolatable helium leak: 28.0 (3), APC4 failure: 8.3 (1), GPC fail-to-synch: 23.7 (3), nonisolatable helium leak: 114 (9) seconds.

Because the total time it took to solve each fault varied considerably, to understand how much of the fault time was spent using the FM display, we normalized the data by computing the percentage of the total fault time that was spent looking at the FM Display. These data are more comparable to each other than total times are. For the four malfunctions, the percentages of the total malfunction time spent looking at the FM Display were: isolatable helium leak: 50 (4) %, GPC fail to synch: 47 (1) %, nonisolatable helium leak: 32 (2) %, and APC4: 38 (3) %.

The eye movement data verified that participants used the FM Display quite extensively. The next question was which aspects of the FM Display were utilized. One possibility is that participants used the display primarily as an input device in which the participant merely hit the "Accept" button. If participants chose to use FAMSS in this way, we would not expect them to access the information available in the schematic section, or even the text, and only focus on the "Accept" button. This would reduce participant's workload, but take them out-of-the-loop and result in low situation awareness. Another possibility is that the participants used the procedure instruction on the FM display as a type of automatic, electronic FDF. In this case, the participant would read the FM procedures, which are similar to the FDF, but not utilize the Fault Management Schematic section (hereafter, referred to as FM Schematic). Finally, participants might have gazed at both FM Schematic and text (hereafter, referred to as FM Text) sections, and (in the case of the helium leaks) used the procedural information embedded in the graphic to better understand the malfunction and the text command. Figure 3.16 resolves these issues by showing that the participants utilized the FM schematics for each of the malfunctions, but more so for the helium faults where the schematics contained the redundant procedures coding.

[2] The APC4 mal required no action. For it, we included the 20 sec of data beginning at its master alarm and ending at the GPC master alarm.

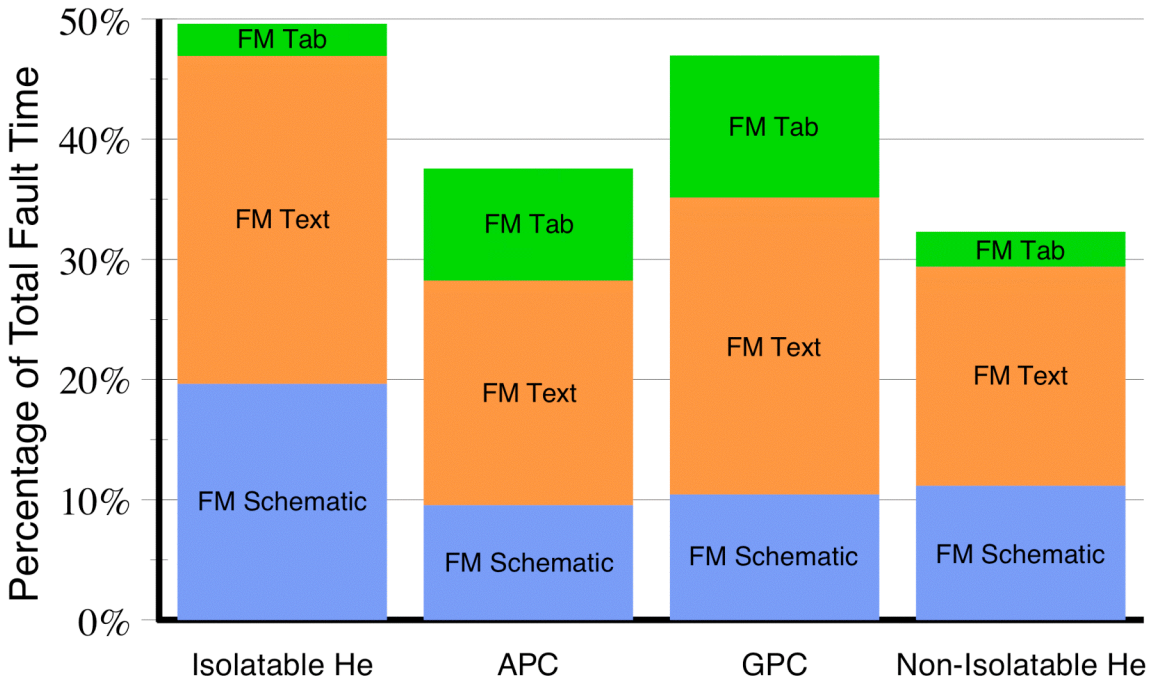


Figure 3.16. *Percentage of total fault management time spent fixating on specific regions of the Fault Management Display for the four malfunctions included in the study. Blue represents the schematic section; orange represents the text section; green represents the malfunction tab. Isolatable He = isolatable helium leak; APC = APC4 subbus failure; GPC = GPC fail to synch; Non-Isolatable He = nonisolatable helium leak.*

The usage of the FM Display, while generally similar for each of the malfunctions, did show some interesting differences. First notice that the total percent time spent looking at the FM display was greatest for the isolatable helium leak and GPC fail to synch malfunctions. Using FAMSS, participants generally focused much of their attention on solving these malfunction quickly. In contrast, for the longer nonisolatable helium leak malfunction, much of the fault time did not require any action and was spent waiting for the countdown indicators to expire, and the participants could utilize much of this time to return to nominal scanning. This issue is discussed further in Sections 1.4.3, 3.4.2, and 3.3.2.2. Also notice that the usage of FM Schematic was greatest for the isolatable helium leak. This may be due to the helium schematics containing the most procedural information (see Section 2.5.2.3.2 for a discussion of differences in the schematics for the various malfunctions and Section 3.4.2.2 for a discussion of how participants utilized the schematic). The percentage of FM Schematic time for the nonisolatable helium leak is lower than that for the isolatable helium leak, perhaps because, for the nonisolatable helium leak, more percentage of the time was spent on nominal scanning, and on the countdown indicators (which were adjacent to the FM text region, and thus were included as FM Text fixations).

Much of the FM schematic information was also available on other displays. Did the participants access these other schematics (see Section 2.4.2), or did they exclusively use the FM

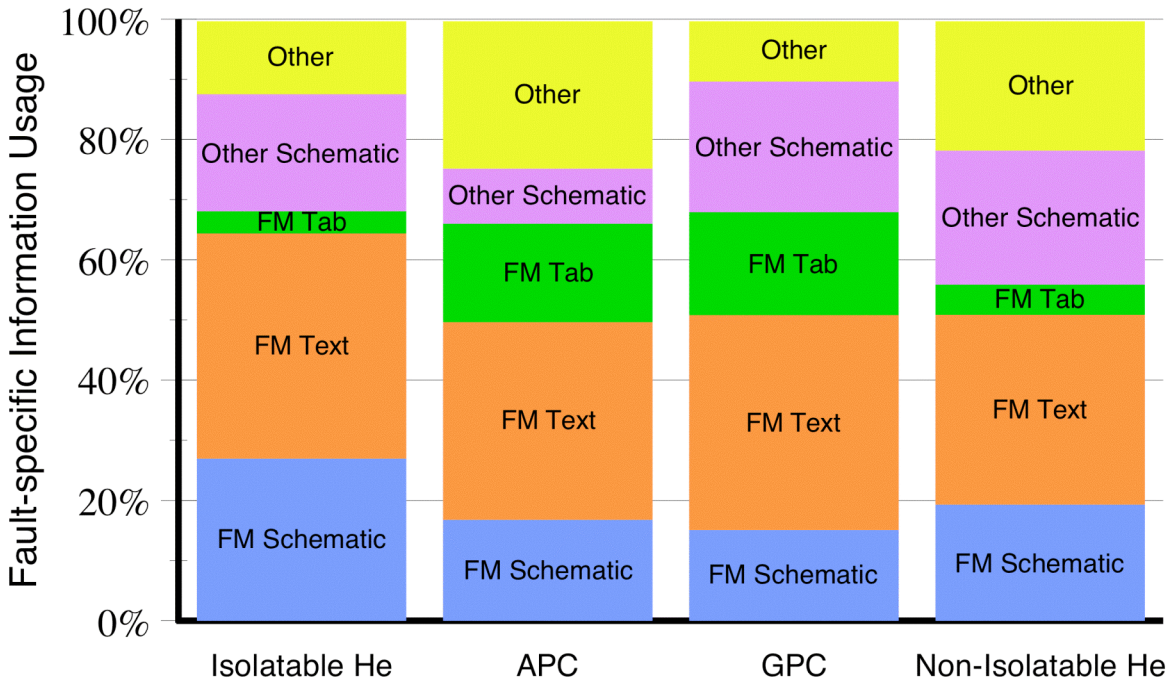


Figure 3.17*. Relative fault-related display usage during the malfunctions. The bars show the relative usage of five different display regions, the three shown in Figure 3.16, FM Schematic, FM Text and FM Tab, plus two additional regions, Other Schematic and Other Fault. The bars were normalized to sum to 100%, so that they show the percentage of time spent on each of these regions relative to the total time spent looking at all fault-related displays. "Other Schematic" = fixation on other displays which contain fault-related schematic information (MPS Sum for the nonisolatable and the isolatable helium malfunctions, EPS for APC4 malfunction, and DPS and the overhead GPC displays for the GPC malfunction). "Other Fault" = other non-schematic areas that may be used during fault management (relevant switch panels, Fault Log, Fault Messages, and Keyboard. For the nonisolatable malfunction, "Other Fault" also includes the ADI/HSI region; for the APC4 malfunction, Helium Schematic and Ullage Pressure regions on MPS SUM).*

text, schematic, and accept button to solve the malfunction, without any cross-checking or verification with the rest of the cockpit? The eye-movement data analysis provides the unique ability to quantitatively answer this question by measuring the amount of time participants spent fixating on non-FM schematics. Participants generally looked at both the FM and other schematics for approximately the same amount of time as shown in Figure 3.17[3], which also shows the percentage of fault-related time that the participants spent looking at other fault-related displays.

3.3.2 Fault Management Display Transition Probabilities

The previous section used the eye-movement data analysis to examine how participants allocated their time to solve the malfunctions. It showed that participants heavily utilized FM Text, FM

[3] For the isolatable helium leak malfunction several of the participants relied nearly exclusively (over 75% of schematic looks) on either the FM schematic (5 participants) or the MPS Sum schematic (4 participants).

Schematic and the other schematic displays, but does not provide much information about how participants coordinated their information acquisition across these regions. Were the Fault Management Display schematics used to supplement the text and improve the operator's understanding of the text command and the underlying physical system's current and commanded states? If so, we would expect frequent eye movements between these two regions. How were the various schematics relevant to the faults, such as the helium system schematics on MPS SUM used in relationship with FM Schematic and FM Text? If the non-FM schematics were primarily used to crosscheck information with FM schematics, we should see frequent eye-movements between the two schematic regions, and not much traffic between FM text and the non-FM schematic region. To address these and similar questions, and to supplement the first-order measurement of duration, we also computed the transition probabilities, a second order parameter which measures the probability of transitioning from one ROI to another.

Our major interest is in understanding the usage patterns of features within the FM display itself, and how participants coordinated their extraction of information from the FM display with information extraction from non-FM schematic regions. Therefore, we examined transitions between four regions, the FM Text, the FM Schematic, the other schematic (see Figure 3.17 for the definition of the other schematic region for the four malfunctions), and all other non-fault-management-related regions, which are denoted as "other" ROI. To understand how frequently participants moved their eyes from one region to another, we computed a separate transition probability matrix for each malfunction. We included only data from participants who successfully completed the malfunction.

The transition probability matrices were defined and computed as described below. Only transitions from one region to another were included: transitions within the same region were neglected (e.g. reading saccades, or consecutive saccades examining different sub-regions within a schematic), so the diagonal terms in the matrices were defined to be zero and are not shown. The rows represent transitions from one region to the other three regions (i.e., the first row is the probability of making a saccade from FM Text to each of the other three regions), so each row by definition must sum to 1.0, while there are no constraints on the columns. Also, the transition probabilities are not directly related to the duration data, because for example, a long duration may correspond to a long time spent within a region and thus a single transition, while many rapid saccades into and out of a region would produce a high transition probability for that region, but a brief duration. The resulting transition probabilities (Figures 3.18-3.21) are plotted as 4x4 matrices with the each cell's transition probability indicated as an orange number and the brightness of the cell's background (the brighter the background, the higher the transition probability)

3.3.2.1 Transition Probabilities for the Isolatable Helium Leak

Figure 3.18 shows the transition probabilities for the isolatable helium leak malfunction. It shows that FM Text acts as a hub for each of the other regions, as the highest transition probability (all about 0.5) is from each region to FM Text. From FM Text the dominant transition is to FM Schematic, as would be expected if it is being used in conjunction with FM Text. The transitions between FM Schematic and MPS SUM Helium Schematic were relatively uncommon, 0.15 and 0.18, so it does not appear that the two schematic regions were directly

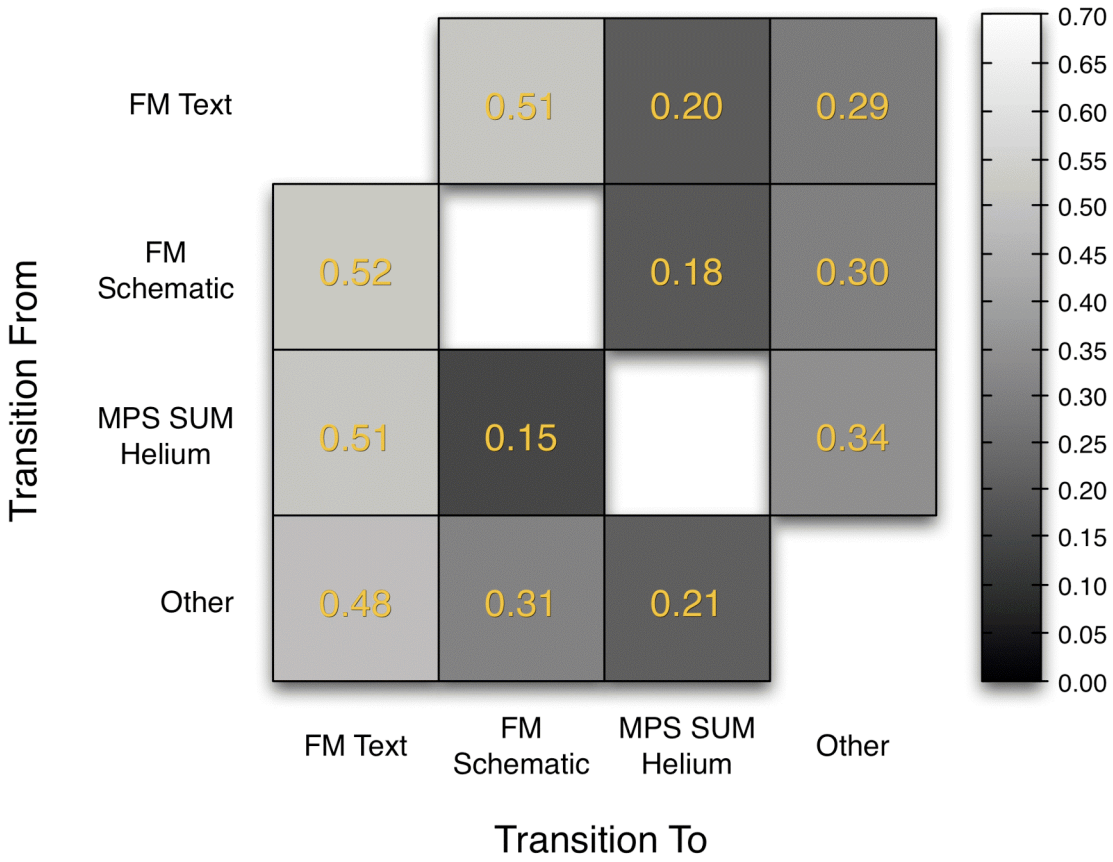


Figure 3.18. *The probability of transitioning between one region of interest (ROI) and another during the isolatable helium malfunction. FM = Fault Management Display. FM Schematic = schematic region of Fault Management Display. MPS SUM Helium = helium supply systems schematic region of CAU MPS SUM display. Other = all other regions of interest.*

crosschecked very often. All regions have an approximately 0.3 chance of transitioning to the "other" ROI.

3.3.2.2 Transition Probabilities for the Nonisolatable Helium Leak Malfunction

Figure 3.19 shows the transition probabilities for the nonisolatable helium leak. This is the long duration fault in which participants must wait for the countdown indicators to expire before opening the interconnect valve and shutting down the engine. One of the design goals for the Fault Management Display was to free up time during this malfunction for the participants to return to nominal monitoring of the other displays. This would be reflected in higher transitions to the "other" ROI. The data show that transitions from MPS SUM Helium Schematic to the "other" ROI nearly doubled compared to the isolatable helium leak (0.64 compared to 0.34); a similar increase occurred from FM Text to the "other" ROI (0.49 compared to 0.29). These are consistent with the participant periodically switching from working the malfunction to performing nominal checks. The transition from FM Schematic to the "other" ROI does not show much increase (0.38 compared to 0.30). This is consistent with FM Schematic being used as it was for the isolatable helium leak, primarily for cross checking with FM Text. Similarly,

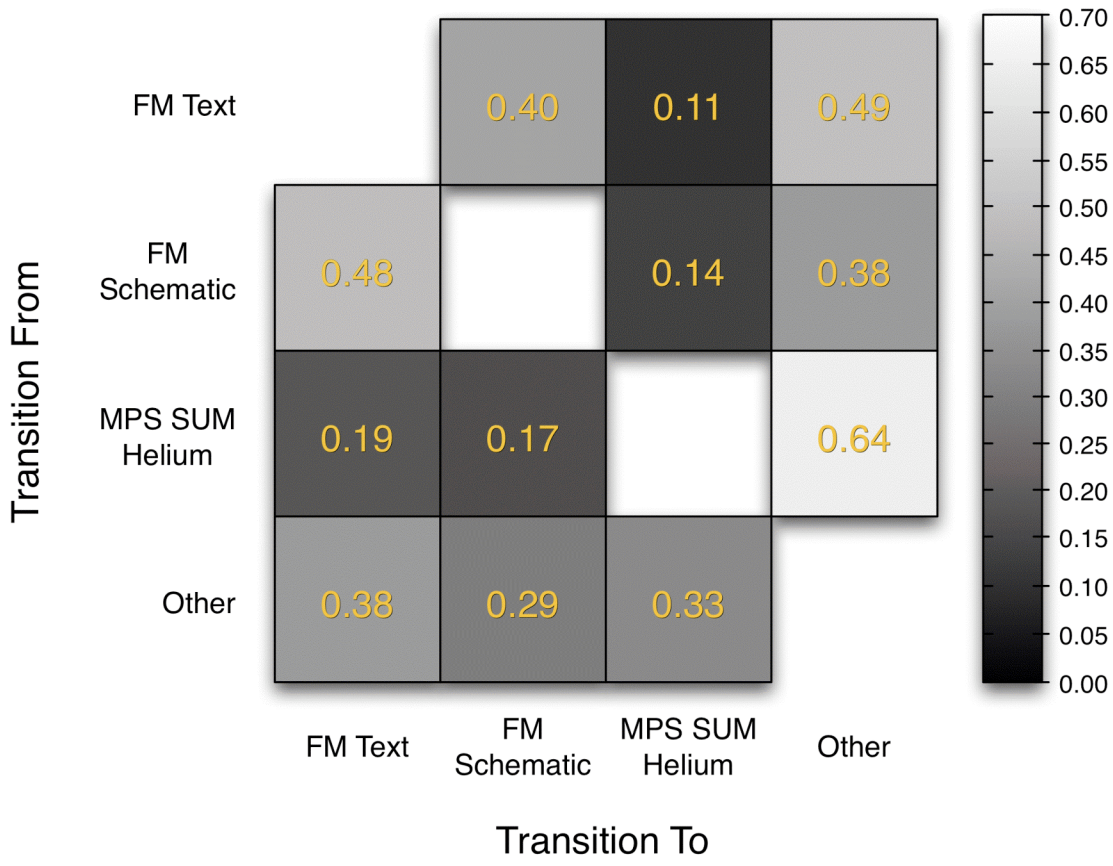


Figure 3.19*. The probability of transitioning between one region of interest and another while working the nonisolatable helium malfunction. FM Text = Text region of Fault Management Display. FM Schematic = helium schematic region of Fault Management Display. MPS SUM Helium = helium supply systems schematic region of CAU MPS SUM display. Other = all other regions of interest.*

for both the isolatable and nonisolatable helium leak, the transitions from FM Text and FM Schematic to each other remain high (0.48 and 0.40), and there is not much traffic between FM Schematic and MPS SUM Helium Schematic (0.14 and 0.17). However a change in behavior is indicated by the large decrease in transitions from the MPS SUM Helium Schematic to FM Text (0.51 for the isolatable leak decreases to .19 for the nonisolatable leak). For the latter, these transitions appear to have mostly shifted to "Other" ROI, perhaps because of increased nominal scanning involving MPS SUM Helium schematic. Further analysis and examination of the transition probabilities for these regions would be needed to validate this hypothesis.

3.3.2.3 Transition Probabilities for GPC Fail To Synch

Figure 3.20 shows the transition probabilities for the GPC fail to synch malfunction. Many transition probabilities are similar to those for the isolatable helium leak (the most closely analogous fault to GPC fail to synch, as both required speeded responding). For example, for both malfunctions, FM Text is the major destination choice from the "other" ROI (0.45 for the

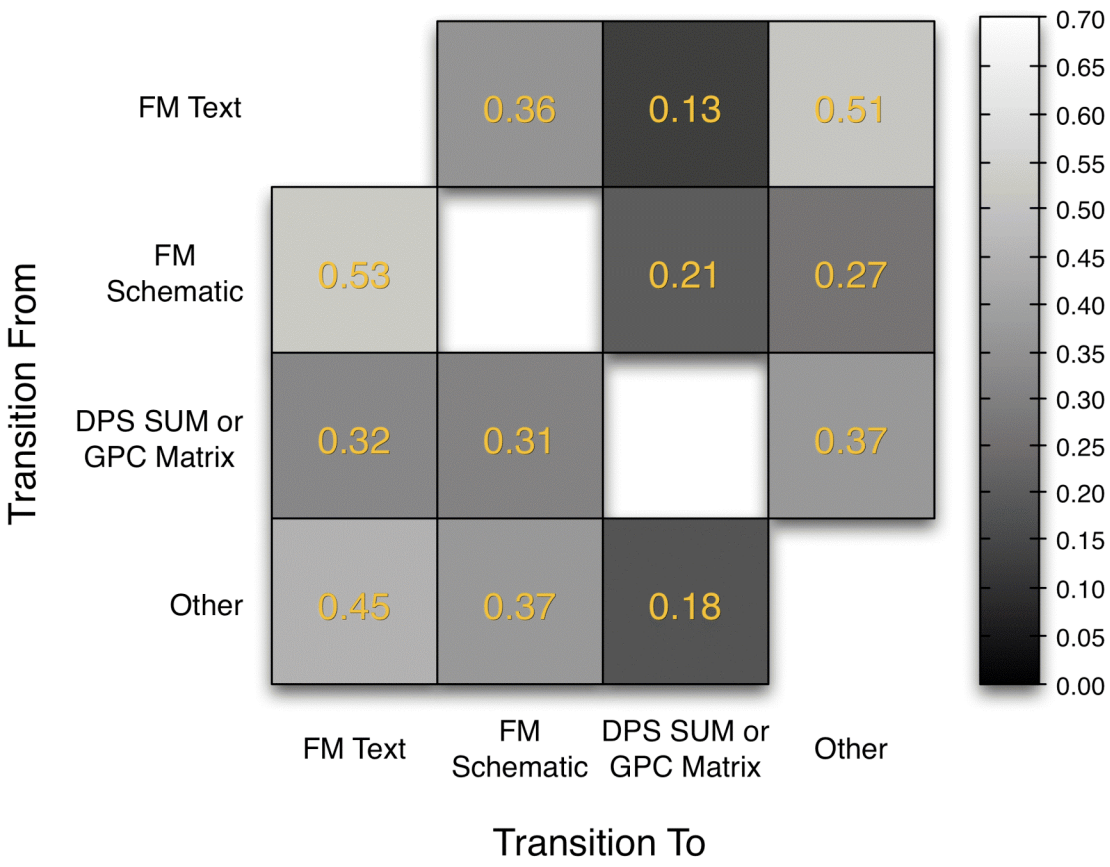


Figure 3.20. *The probability of transitioning between one region of interest and another during the GPC fail to synch malfunction. FM Text = Text region of Fault Management Display. FM Schematic = GPC schematic region of Fault Management Display. DPS SUM or GPC MATRIX = CAU DPS SUM display or GPC Matrix on overhead panel. Other = all other regions of interest .*

GPC fail-to-sync compared to 0.48 for the isolatable helium leak), consistent with FM Text maintaining its role as a hub. However, the top row reveals interesting systematic discrepancies between malfunctions. In the GPC case, transition probabilities between FM Text and FM schematic were significantly lower than for the isolatable helium leak (0.39 compared to 0.51; t[9] =2.4, $p < 0.05$). Transition probabilities between FM Text and "Other" were significantly higher than for the isolatable helium leak (0.51 compared to 0.29; t[9] = 4.0, $p < .01$).

In general, this pattern indicates that FM Text was less tightly coupled to FM schematic for the GPC fail-to-synch malfunction. In particular, participants were less likely to transition directly from FM text to FM schematic. We speculate that this is due to the fact that procedures were not redundantly coded inside FM Schematic (see Section 2.5.2.3.2 for a discussion of this). With less information overlap between that section and the text, there was less utility in the crosscheck.

3.3.2.4 Transition Probabilities for the APC4 Malfunction

Transition probabilities for the APC4 subbus failure, shown in Figure 3.21, are considerably different than the other malfunctions. To understand these differences, it is important to note that most participants did not navigate to the EPS SUM display, which contained information about the APC4 failure in graphic form (this is consistent with the small amount of time spent on this region: see the "Other Schematic" slice in Figure 3.17). A large part of the reason for this may simply be that the graphical representation of the problem on EPS SUM was more-or-less replicated in FM Schematic (see Figure 2.5), which was more readily accessible. This could account for the very low transition probabilities to EPS SUM. However, those participants who did bring up EPS SUM tended to transition from EPS SUM to "Other" far more than to FM Text, in sharp contrast to the pattern we see in the other malfunctions, where the strong tendency was to transition from other cockpit regions containing fault-related schematic information to FM Text. Similarly, there is much less traffic from FM Schematic to FM Text than in other malfunctions; instead, again, most transitions are from FM schematic to "Other."

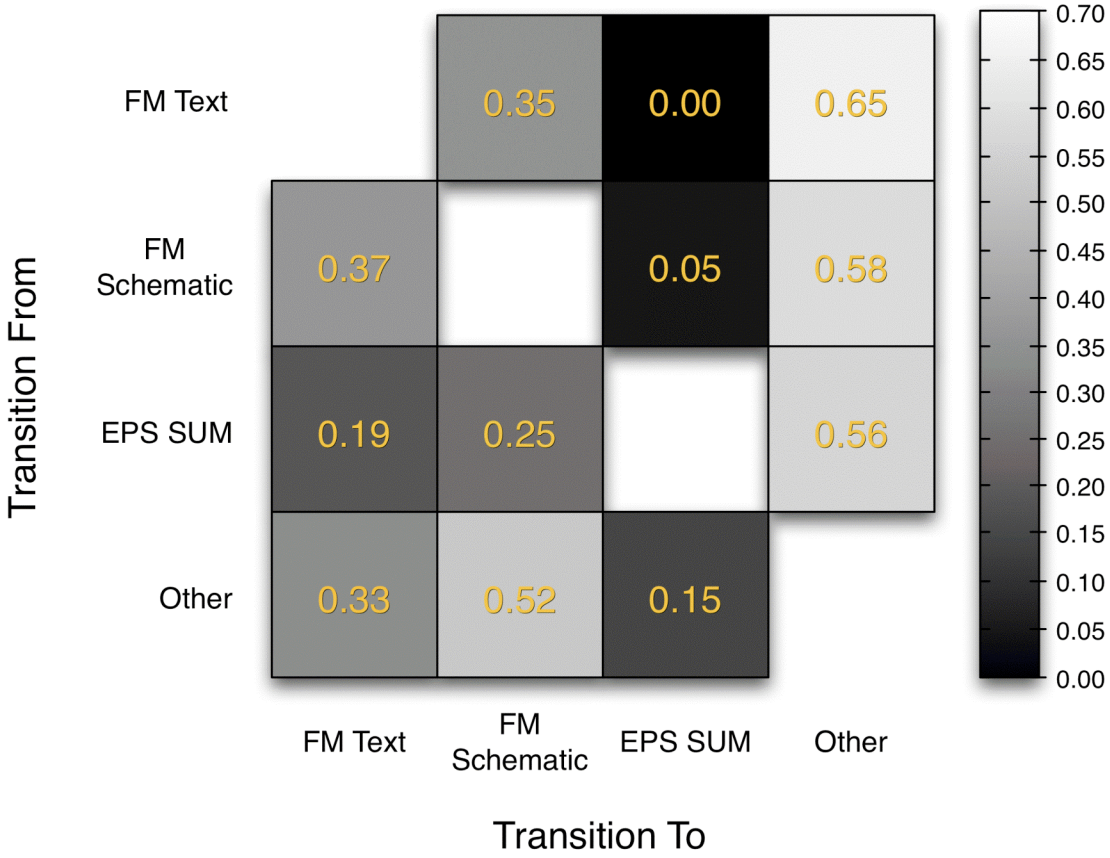


Figure 3.21. *Probability of transitioning between one region of interest and another during the APC4 subbus failure (prior to GPC fail to synch only). FM Text = text region of Fault Management Display. FM Schematic = schematic region of Fault Management Display. EPS SUM = CAU EPS SUM display. Other = all other regions of interest.*

These patterns are consistent with the conclusion that FM Text was not used as a hub as much as for the other malfunctions. In fact, some of the hub role appeared to shift to FM Schematic (as reflected by the high [0.52] transition probability from "other" to FM schematic). Perhaps this is because, for this malfunction only, FM Text did not contain an actual procedure, so there was less need to continually reprocess it and crosscheck with schematic information.

The other outstanding issue is the dramatic increase in transitions to the "Other" ROI compared to the other malfunctions. This is consistent with participants checking other displays to examine the cockpit signatures associated with the "daughter" consequences of the fault.

In summary, transition probabilities provide considerable information about how participants used the FM display to solve the malfunctions. Each malfunction produced different patterns in the transition probabilities. These differences in the transition probabilities show that participants adopted different strategies for each malfunction, which depend on the type of malfunction, the time constraints and the type of FM schematic. These data provide information uniquely available from eye movements, important for the evaluation of proposed interfaces and of user strategies and abilities.

3.4 Subjective and Objective Evaluation Tools: Making the Connection

Ratings of situation awareness and workload, collected after each run, showed that participants rated their situation awareness higher, and their workload lower, in the FAMSS Condition than in the Baseline Condition (see sections 3.2 and 3.3). However, we also collected an extensive set of individual measures of performance on each run, such as malfunction resolution accuracy, resolution response time, and duration of gaze at regions such as the AESP FDF. It is of interest to determine whether workload and situation awareness ratings were sensitive to any of these performance-based measures, in part to determine the proper role for subjective ratings in human factors evaluations of VSE operations concepts. In this section we ask whether run-specific variance in rated workload and rated situation awareness bears any relation to objective measures of performance such as malfunction resolution accuracy, resolution latency, and time on task as measured by eye fixations.

Response Accuracy and Workload. For the isolatable helium leak (single malfunction run) in the Baseline condition, eight participants were successful and six were unsuccessful. In the other three combinations of runs with conditions (single malfunction with Baseline, single malfunction with FAMSS, multiple malfunction with FAMSS), resolution accuracy was much more one-sided. The more even split for the single malfunction run in the Baseline condition provides a convenient basis for determining the possible correlation of response accuracy with workload. In fact, Bedford workload ratings were nearly identical for the two groups (3 out of 10 for the successful group and 4 out of 10 for the unsuccessful group). TLX workload ratings were identical (4 out of 10 for both groups). Even the TLX component of frustration (which might be expected to be most sensitive to success on the task) was identical (4 out of 10 for both groups). T-tests showed no statistical difference in those subjective differences between the two groups for any of those three workload metrics.

An alternative means of assessing possible correlations between accuracy and workload is by using the total number of correct switch throws as a measurement of accuracy. For the multiple malfunction condition in the Baseline condition, the number of correct switch throws ranged from 0 to 5. However, no statistically significant correlation between workload and that measurement of accuracy was found.

Response Time and Situation Awareness. For the single malfunction run, the eight participants who completed all steps in the Baseline condition also completed those steps correctly in the FAMSS condition. Their response times for completing all steps were correlated with their subjective situation awareness scores (for diagnosing the malfunction and also working the malfunction). The response times ranged from 77 to 335 seconds (Baseline) and 28 to 50 seconds (FAMSS). Situation awareness ranged from 3 to 10 (Baseline) and 8 to 10 (FAMSS). The correlation coefficient between Baseline response time and situation awareness ratings for diagnosing the malfunction was a relatively low 0.3. Similarly, the correlation between Baseline response time and situation awareness ratings for working the malfunction was also 0.3. Neither correlation coefficient was statistically significant, indicating that subjective situation awareness ratings are not sufficient in themselves to provide direct insight into performance. The correlation coefficient between FAMSS response times and subjective situation awareness was higher (0.6 for diagnosing the malfunction and 0.4 for working the malfunction). As with the Baseline condition, neither correlation coefficient was significant, possibly due to the narrow range of situation awareness ratings in FAMSS

Response Time and Workload. For the single malfunction run, correlations of response time with workload were evaluated for the eight participants who completed the procedures correctly in both conditions (Baseline and FAMSS). A priori, it would seem plausible that rated workload would be sensitive to the amount of time spent working a malfunction. Indeed, for the FAMSS condition, the correlation of 0.8 is statistically significant, $F(1,6)=13.9$, $p < 0.01$. However, for the Baseline condition the low correlation of 0.1 was not significant.

Eye Fixations (on FDF) and Workload. For the single malfunction run, eye movements of 11 participants were recorded accurately enough to identify across the entire run (approximately 8 minutes, 24 seconds) whether they were fixating on the AESP FDF or other regions of the cockpit. One hypothesis is that the percentage of time spent on FDF represents a high workload period, since we know from numerous other measures that FDF navigation was a difficult and time consuming operation. For the Baseline condition, the amount of time spent looking at FDF ranged from 16 to 138 seconds (depending on participant). For the FAMSS condition, the range was much smaller (0 to 14 seconds), since virtually all participants eschewed the use of paper in favor of the Fault Management Display. The correlation coefficient between Baseline FDF time and Bedford workload was small (0.1) and not statistically significant. Similarly, the correlation coefficient between FAMSS FDF time and Bedford workload was also small (0.3) and not statistically significant.

Eye Fixations ("Off-Task") and Workload. An alternative hypothesis is that the percentage of time spent "Off-Task" (meaning fixating nominal regions of the cockpit) should be inversely correlated with workload. However, once again, no statistically significant correlation was found.

Use of individual display elements and usability ratings. Another example of a dissociation between subjective measures and objective performance comes from comparing eye movement data, revealing how frequently participants actually looked at a display feature, compared to how valuably they rated the feature on the usability scales. The particular example comes from participants' ratings of the usefulness of the helium schematic on the FM display. All participants rated the usefulness of the schematic between 7 and 10 on a 1-10 scale. Most of our participants looked at the schematic frequently, usually as part of a crosscheck with the text section. However, for three of our participants, the number of fixation durations on the fault management schematic was almost zero; these participants preferred to crosscheck the text-based instructions with the helium schematic on MPS SUM. Nevertheless, their ratings of the usefulness of the fault management schematic were as high as all the others.

Eye Fixations (Nominal runs) and Workload. Recall that we found a significant difference in rated workload on the nominal runs between Baseline and FAMSS days, a somewhat puzzling result since the information acquisition and processing requirements did not differ. A final set of analyses was performed on just the nominal runs to determine the potential relationship between the differences in rated workload and the total number of individual eye fixations. The difference in total number of individual eye fixations for each participant across Baseline and FAMSS was correlated with their rated difference in TLX workload on Baseline and FAMSS testing days (based on the average rating for the two nominal runs on each day). The correlation coefficient was low (0.2) and not significant. Similar results were found for potential correlations for total eye transitions and Bedford workload.

The results in this section indicate that, in general, participants' workload and situation awareness ratings were surprisingly insensitive to specific aspects of their performance. Possible explanations are simply that a participant's subjective understanding of the environment may not translate into strong performance. This result is in line with previous studies (such as Endsley, 1995), which found that performance can be affected by a number of factors that might not be related to situation awareness. Similarly, one participant might work diligently at performing a task (producing a high workload), and still fail at that task. Another might exert an equally strong effort and succeed at the task. Clearly, an evaluation in a complex setting such as a spacecraft cockpit requires numerous evaluation metrics, not only subjective metrics like workload ratings but also objective metrics such as performance and eye tracking.

4 Discussion

Our discussion of the FAMSS evaluation is organized around the following topics: Section 4.1 covers the quantitative benefits of FAMSS, and where the biggest improvements occurred. Section 4.2 discusses what the evaluation revealed about FAMSS design and interface shortcomings, and possible design modifications to overcome those shortcomings. We also discuss what sorts of additional functionality FAMSS might require in order to handle malfunction scenarios where conditions are not as "cut and dried" as in the evaluation, and there is some ambiguity as to root-cause determinations, for example. Section 4.3 discusses the implications of our results for more "incremental" changes to current fault management

operations that could be applied to the first block of next-generation spacecraft design, where onboard avionics capabilities may fall short of what is needed to support full-blown FAMSS. Section 4.4 considers what our integrated approach to evaluation and testing metrics told us about the comparative strengths and weaknesses of the evaluation tools and testing metrics. Finally, Section 4.5 discusses possible directions for future work on developing and evaluating advanced displays, operational concepts, and human-computer interaction for next-generation spacecraft cockpits.

4.1 FAMSS: Expected and Actual Benefits

A central goal of the evaluation was to identify and quantify the benefits of an integrated concept for human-machine partnering for real-time fault management. We can best characterize these benefits by first identifying what our evaluation revealed about the source of fault management difficulty without FAMSS, that is, in our CAU baseline condition. We will identify and highlight sources of difficulty via a discussion of results obtained with those traditional behavioral measures, accuracy and latency.

4.1.1 Errors

We begin this section by pointing out and emphasizing the extensive training our participants received in the constituent malfunctions, their cockpit signatures, and how to manage them correctly. Training on similar malfunctions, across various shuttle cockpit display suites, went back several years, and very focused and extensive training and coaching in the CAU condition occurred as recently as the morning of the day the data was collected. If anything, our concern at the outset of the study was that we were guilty of overtraining, potentially reducing the sensitivity of the study to any FAMSS benefits.

As the study revealed, nothing could have been further from the truth. Fault-management errors were sufficiently common that only 43% of the off-nominal scenarios were resolved correctly. On the one hand, this result testifies to the extreme difficulty of fault management operations on today's spacecraft, and why it takes two years of astronaut training and continuous practical experience in ground simulators to attain and maintain operational proficiency. But that very proficiency can work against human factors researchers who are interested in identifying sources of difficulty and user interface problems with a cockpit. Using less well trained participants, such as those in our study, enables these "latent" problems to manifest themselves in measurable behaviors such as errors. Analysis of these overt behavior can help determine priorities for improving operations on VSE vehicles.

Which brings us back to our errors. In the baseline condition, where most of the errors occurred, our participants were responsible for making correct root-cause determinations of clusters of C&W events, reading and navigating through the flight data files, and manually throwing switches. These activities are all potential sources of error. Our study revealed errors with each of the following:

Determining root cause. In the Baseline Condition, two participants failed to recognize that the APC4 subbus failure was behind the closed isolation valve on MPS SUM and the off-nominal

indications on the CAU APU/HYD SUM display. These individuals attempted to work these "daughter" problems as if they were bona fide systems malfunctions.

Navigating the AESP FDF. While working the isolatable helium leak, four participants made a serious error of procedural omission (not taking the helium interconnect to IN-OPEN after isolating the helium leak) reflecting a failure to navigate through the FDF after successfully completing the procedures to isolate the leak to Leg B. Specifically, these participants failed to navigate from the isolatable section of the MPS He checklist to Step 10 (Figure 1.4). Indeed, the results of several analyses converged on the conclusion that FDF navigation poses a particularly large challenge to effective fault management, a point we shall return to shortly.

Another form of FDF navigation error occurs when an operator follows an incorrect path through the checklist because of a lack of situation awareness of system state and status. This error occurred repeatedly on the more complex run. As previously noted, the section of the AESP FDF devoted to APC subbus malfunctions contains a prospective memory instruction not to follow the normal path through the MPS He P checklist and try to isolate a leak in the affected helium supply system, should the symptoms of a leak appear later in flight. When the helium leak actually occurred, five participants followed the incorrect path through the checklist, closing both helium supply legs and shutting the engine down (two additional participants started down the incorrect path, but halted before closing both isolation valves). As noted in the results section, verbal callouts on the videotapes revealed that many of these participants understood the significance of the APC4 failure at the time it occurred, (30 sec into flight) with verbal confirmation that they should not attempt to isolate a helium problem. However, following a very busy period of interpolated activity, they failed to retrieve and act on this knowledge when the critical point in the flight arrived. Further supporting the conclusion that this was a memory failure, as opposed to a lack of understanding of the earlier APC4 failure and its implications, every one of these participants chastised themselves prodigiously, sometimes in rather colorful language, right after the engine shutdown, for making the procedural error. Clearly, they had encoded the knowledge not to take the action, but some combination of forgetting and "habit capture", the tendency to revert to well-learned and practiced activities when distracted and under high workload, rendered that knowledge inaccessible.

Motor errors. Several instances were recorded where participants intended to toggle a particular switch to an intended new position, but failed in the attempt.

In the FAMSS condition, most of the errors in these categories were eliminated. The source of the elimination was sometimes obvious, as with eliminating slips and other low-level motor errors by automating switch throws. In other cases, the source was less intuitively obvious. For example, the benefits of eliminating the need to retrieve fault management instructions from prospective memory in order to navigate a FDF checklist accurately would not have been nearly as evident without the results from the Baseline condition.

4.1.2 Malfunction Resolution Times

FAMSS reduced the time needed to complete fault management activities by as much as two thirds of the time required in the Baseline condition. In a dynamic flight phase, where malfunction severity and mission impact often grow over time (for example, leaks often grow in

size), time savings of these magnitudes are a highly attractive component of the FAMSS package, just from the perspective of reducing the risk of the malfunctions themselves.

But, again, direct comparison of performance between FAMSS and Baseline conditions revealed that the time savings associated with FAMSS did more than reduce the risk posed by potential growth in malfunction impact over time. Reduced resolution times also freed up mental resources to deal with subsequent malfunctions in a more efficient manner. Again, this point would not be nearly as obvious without the Baseline results, with their dramatic confirmation that, when working malfunctions with today's cockpit interfaces, malfunction-handling efficiency is greatly impacted when the malfunction has to be time-shared with another problem. Analysis of eye movements revealed that little or no parallel processing was possible between APC4-related activities and GPC fail to synch activities, consistent with the fact that most of the constituent activities (FDF navigation, locating and throwing switches, examining particular regions of interest on cockpit displays) require focused attention. However, if the only disruption to task sharing was simply that people can't do "two things at the same time", we would have expected the slope relating the duration of temporal overlap between APC4-related processing and the time to initiate the first procedure for the GPC4 failure to be exactly 1.0. That is, for every additional second of APC4 activity, the time to initiate the first GPC fail to synch activity would be delayed by exactly one second. The actual slope of the regression line, however, was closer to 1.5 sec, meaning that every second of additional overlap had a greater than one second impact on time of completion of the first GPC4-related FDF procedure. The slope suggests that working a task as difficult as fault management on the shuttle has a distinct mental task management component, perhaps related to the need to organize, schedule, and coordinate so many disparate activities and forms of vehicle and systems knowledge. Task management-related activities may themselves be quite fragile and subject to disruption when they have to be time-shared with another fault management task.

In the FAMSS cockpit, of course, the mean length of time spent on the APC4 malfunction was dramatically lower than in the Baseline Condition, so there was dramatically less time-sharing with the GPC problem. In turn, this drastically reduced the variance in time to initiate the GPC failure activities, consistent with the discovery that interference from the APC4 malfunction in Baseline was virtually the sole source of inter-participant variance in time to begin working the GPC4 problem.

4.1.3 Workload

Another benefit of FAMSS was lower workload ratings. Workload is commonly associated with the amount of spare capacity a crewmember has to deal with additional problems or demands. Currently, the nearly constant contact that crews have with the ground can help alleviate fault management and other sources of operations-related workload through direct ground support. However, once VSE vehicles leave the vicinity of LEO, ground support for vehicle health management will transition from (virtually) real time, to near-real time, to completely unavailable (depending on vehicle distance and the temporal severity of the malfunction). VSE vehicles will have to operate in a more autonomous fashion than today's spacecraft. A system like FAMSS would give crews more ability to "stay ahead of the vehicle" and work spacecraft operations with less real-time assistance from the ground.

4.2 FAMSS: Lessons Learned

Within the confines of this particular study, FAMSS did not exhibit many deficits. There are many broader operational situations that FAMSS could not support in its present stage of development, a topic we will return to shortly. Meanwhile, two specific problems with the FAMSS interface were revealed in the multiple-malfunction scenario. The number of incorrectly resolved scenarios increased to nearly 70% in the Baseline condition but also increased in the FAMSS condition, where nearly 20% of participants were unable to manage all three malfunctions accurately. Most of the difficulty was with handling the GPC malfunction.

The first malfunction in the multiple-malfunction scenario was to an electrical subbus. Again, this failure had no associated procedural actions, and was introduced to increase the complexity of the scenario and impose a prospective memory requirement on participants. The second malfunction (GPC fail to synch) was, of course, entirely independent of the APC4 failure; it was not a propagated – so-called "daughter" – failure. Nevertheless, the participants in the FAMSS condition who failed to work the GPC problem misinterpreted it as a daughter problem of the APC4 problem, and decided they did not need to work it. This misinterpretation could be just due to lack of training with the FAMSS interface, of course. Further experience with FAMSS might result in a better understanding that because FAMSS only allows root-cause failures to show up on the Fault Management Display (utilizing its ECW-style filtering capabilities), the appearance of a tab on the FM display is an unambiguous cue that a malfunction is present and requires the crew's attention.

Nevertheless, the Fault Management Display itself provided little perceptual support to help establish the independent nature of the GPC failure and communicate to the operator that procedures were waiting to be approved. Although a GPC tab appeared on the right side of the display, the fault management window continued to show the APC4 display (despite the fact the APC4 had no overt actions to complete and GPC4 did). Perhaps FAMSS should be designed with some simple prioritization rules, one of which would be that if multiple malfunctions are currently "active" (i.e., unresolved), and the current fault management page is occupied by a malfunction that has no overt procedures, that page should be replaced automatically.

More generally, the FAMSS concept for handling multiple simultaneous malfunctions needs further development. For example, the current approach of displaying the tab of a malfunctioning system needs to be extended to deal with multiple independent malfunctions in the same system. If three EPS malfunctions occur, how should the crew be made aware of the fact that there are multiple problems to handle? How should the crew navigate between the three fault management pages? Would repeated presses of a single EPS tab be the optimal navigation interface, or should three independent EPS tabs appear? These and related issues require further study.

Improve Cockpit Signatures of Downstream Failures. One pattern that was quite clear in the eye movement data was that participants' attention was constantly drawn to the cockpit signatures associated with the consequences of the APC4 malfunction, even in the FAMSS cockpit where the root cause was clearly annunciated. The distracting power of these signatures might be reduced if FAMSS provided explicit information about the consequences (impacts) of systems

malfunctions, and by coding the cockpit indications in a unique manner. For example, similar to the way missing values are handled in CAU (color-coded cyan), we may need to color-code propagated failure signatures to distinguish them from root-cause signatures.

4.3 Incremental Improvements to Fault Management: "FAMSS Lite"

The level of avionics integration and data sharing necessary to support full-blown FAMSS may not be available on the initial build of the next generation space vehicle. The functionality required may only emerge through incremental upgrades to vehicle capabilities in much the same fashion as occurred with the shuttles. What do our results have to tell us about targeted improvements that could be made to today's fault management interfaces that would have the biggest impact on fault management efficiency?

Several results from the Baseline condition converge on the conclusion that navigating through the paper flight data files contributed disproportionately to the overall difficulty of fault management operations. From our accuracy analysis, we found that while navigating the isolatable helium leak section of the AESP FDF, four participants failed to proceed from Step 6 (CLOSE ISOL B) to Step 10 (take the interconnect to IN-OPEN). Participants who did make it to Step 10 had long and highly variable navigation times, some because they "left the book" entirely. A combination of videotape viewing and analysis of the durations of fixations on the relevant sections of the AESP FDF revealed that even participants who did work the checklist in a relatively uninterrupted fashion took, on average, approximately 16 seconds to navigate from Step 6 through Step 10. This compares to a predicted navigation time of only 8 seconds derived from summing the relevant behavioral primitives included in the APEX-GOMS model of isolatable helium leak behaviors (Appendix B).

The source of the difficulty is almost certainly the structure and physical layout of the MPS He P section. Specifically, looking again at Figure 1.4, the procedures for an isolatable helium leak are distributed across multiple distinct subsections defined by horizontal lines that function as effective perceptual dividers. In the AESP MPS He P section, Step 6 is separated from Step 10 by two such subsections. Thus, if anything, the physical layout and structure of the section are working against the desired behavior, which is to A) ensure that navigation from Step 6 to Step 10 occurs, and B) ensure that the transit from one step to the other occurs in as short a time as possible.

Airline industry employees with experience in designing and developing electronic checklists will not find these results surprising. Although paper checklists have some benefits over electronic checklists in terms of their portability and reliability, their drawbacks can be significant. Numerous error modes associated with paper checklists were recently summarized by Boorman (2000), and one prominent mode is simply "One or more items are skipped in the checklist". As it might on a spacecraft, failing to navigate paper checklists correctly on aircraft have had critical consequences. For example, the National Transportation Safety Board (NTSB) found that the 1987 crash of an MD-82 (killing 156 people) was caused by the failure of the crew to accomplish all the steps in their taxi checklist (NTSB, 1989).

Developing electronic versions of procedure checklists for next generation spacecraft offers numerous opportunities to eliminate the sources of checklist navigation difficulty revealed in our study. Consider the MPS He P section in the AESP FDF, where intervening instructions and subsections separate procedures belonging to the "isolatable leak" navigation path. A perceptual connector, perhaps something as simple as an arrow (perhaps with a dotted rather than a continuous line, to indicate that the navigation path is not direct, and that there are intervening conditionals that must be assessed) or a text-based reminder, could be of significant benefit even in a very low-tech cockpit (i.e., in redesigned paper version). Airline cockpits contain many examples of more advanced navigation aids on their electronic checklists, such as surrounding the currently commanded procedure with an outline box, and automatically moving the box to the next procedure in the appropriate navigation path when the currently selected procedure is completed. These kinds of aids, of course, would also be very effective.

4.4 Evaluation Tools and Techniques

We employed a variety of human performance measurement techniques and evaluation tools to evaluate the FAMSS concept. In addition to the standard suite of human factors evaluation tools (objective performance measurements and questionnaires), we incorporated two less frequently used methods, eye movement analyses and predictive human performance modeling. In this section, we consider the "value added" of these techniques, and well as what we gained by our integrative approach to tool utilization.

4.4.1 Situation Awareness and Workload

Both the Bedford Scale and the NASA Task Load Index showed that FAMSS significantly decreases the workload required to work malfunctions in the shuttle cockpit. An interesting question is how strongly we should rely on these measures when evaluating operations concepts for VSE vehicles. Our results provide several forms of evidence suggesting that subjective ratings are less sensitive to some influences than we would like them to be, and more sensitive to other influences than we would like them to be. Specifically, while these ratings yielded significant differences between FAMSS and Baseline conditions, they were surprisingly insensitive to within-run aspects of participant performance, such as malfunction resolution latency and accuracy. Meanwhile, we cannot rule out the possibility that participants might have provided ratings that they thought we wanted in order to help FAMSS gain traction and acceptance. With only fourteen subjects, all of whom spent many hours associated with this project, such a bias would not be surprising.

Meanwhile, the workload measurements for nominal runs in the Baseline cockpit were rated as significantly higher than workload in FAMSS. In nominal conditions, the participants were, for the most part, required only to monitor the displays. There were no FDF procedures to find, nor were malfunction-related switch throws required. This finding might therefore reflect a context or anchor effect, rather than a true reflection of run-specific workload.

The message is that, although subjective measures such as workload and situation awareness are sensitive to high-level cockpit changes, such as the difference between FAMSS and Baseline

Cockpit conditions, these measures are remarkably insensitive to the particulars of a participants experience on a particular run, such as his eye movement behavior, task accuracy, or time required to complete a task. Clearly, any evaluation that is based solely on subjective measures provides an incomplete (though not necessarily totally misleading) picture of human performance in these complex operating environments.

4.4.2 Eye Movements

Eye movement data augmented the traditional forms of operational concept and cockpit display evaluation in several important respects. In several cases, we were able to start with a behavioral result from one of the more traditional measure of performance, such as the very long and variable latencies to complete the final procedure for the isolatable helium leak, to flag a particularly difficult aspect of a fault management operation. Eye movement analyses then allowed us to burrow into the problem to more precisely determine the source of the difficulty (in this case, FDF navigation). As we have seen, the more precisely we can determine the source of the problem, the more precisely we can generate design solutions.

In many other cases, we were able to propose specific hypotheses about display usage and information acquisition and coordination, and then evaluate the hypothesis via eye movement analyses:

1. Understanding FM Display usage. While traditional measures of performance showed large benefits of the FAMSS system, they were silent on the issue of how participants used the information on the Fault Management Displays to achieve these benefits. Reductions in fault resolution times compared to Baseline could have occurred, for example, if participants simply used the display as a fast input device, ignoring all the fault-related information and just hitting the "accept" button. That and other hypotheses were falsified by eye movement evidence showing that the text-based instructions formed a processing hub, and participants transitioned frequently from the text to the schematic of the display and back.

The high incidence of these crosschecks is consistent with the further hypothesis that when procedures are represented in both text and graphical form, crosschecking the two formats promotes a high level of understanding of FAMMS intentions and recommendations (i.e., it helps operators stay in synch with the automation). Alternatively, the high incidence of crosschecking could be due to the fact that schematics provide a closer match with peoples' mental representations (knowledge) of complex engineering systems, and the functional relations between their operational components, than text. When a procedure calls for a mode reconfiguration, the procedure is validated and verified more efficiently if a crosscheck is performed between the text and the adjacent schematic representation. For the same reason, the crosscheck may facilitate and support rapid determination of system state and status after a reconfiguration has been performed.

2. Comparing actual versus intended usage of display features. Some malfunctions, such as the nonisolatable helium leak, involve deferred procedures triggered at some time in the future when a system or flight parameter passes a predetermined threshold. We designed the deferred procedures version of FAMSS to provide "at a glance" information about the current status of the relevant parameters and their status with respect to these thresholds. The hypothesis was that providing information in this form would facilitate participants' ability to interleave "status

checks" of these parameters and current systems operations on the Fault Management Display with more general information acquisition activities, possibly up to and including the full nominal scan, across the rest of the cockpit.

Eye movements provided a direct way to assess whether these design goals were achieved. In fact, the evidence was not supportive. On the contrary, the results suggested that the Fault Management Display acted as something of an attentional attractor, such that participants tunneled on the display for a longer period of time than they looked at fault-related sources of information in the Baseline condition. Armed with this information, we could consider some mitigation strategies, such as changing the design of the deferred procedure display to discourage tunneling, or performing a further study to determine if additional training and familiarity with the display features would reduce it.

Again, these direct insights into display usage and information acquisition dependencies across displays and their individual features provide valuable guidelines for designers of displays and operations concepts in VSE vehicles. For example, assuming that our discovery that text acts as a hub during fault management operations has sufficient generality (i.e., that the pattern would be seen in other populations, evaluations, and display designs), the result suggests that the display should include additional features, such as oversizing the text or highlighting it in some other way so that the text region is able to attract attention and generate accurate saccades from other displays.

4.4.3 Human Performance Modeling

Predictive modeling methods have been developed to enable evaluation of different display formats without doing expensive human-in-the-loop (HIL) experiments. In theory, high-fidelity models – improved over time by incorporating data from many HIL experiments – enable accurate prediction of the time it takes to perform a task. The results of this study show that human-performance modeling (HPM) can predict response time performance to some extent. However, the steps included in the modeling (Appendix A) assumed an ideal observer, with no interruptions, crosschecks, forgetting, or other disruptions to which actual people are vulnerable. As a consequence, the model greatly underpredicted response times for both the isolatable and GPC fail to synch malfunctions, particularly for the initial and (for the isolatable helium leak) final procedures. The model also underpredicted FAMSS benefits by a considerable margin.

Because these underpredictions were based on an ideal observer, the model predictions may well be more accurate for a population, such as astronauts, that is more highly trained in the spacecraft cockpit than our participants. Thus, the model represents a sort of a boundary condition, placing a lower limit on FAMSS benefits with respect to shortening malfunction resolution times.

We also may be able to "bookend" our estimate for FAMSS benefits for an astronaut population with an upper limit, as well. Figure 4.1 graphs the same conditions as Figure 3.9 including only our fastest Baseline participant (who was also the second fastest participant to complete all procedures in the FAMSS condition), arguably the participant who came closest to astronaut performance. In Baseline, this individual completed the isolatable helium leak procedures in 76

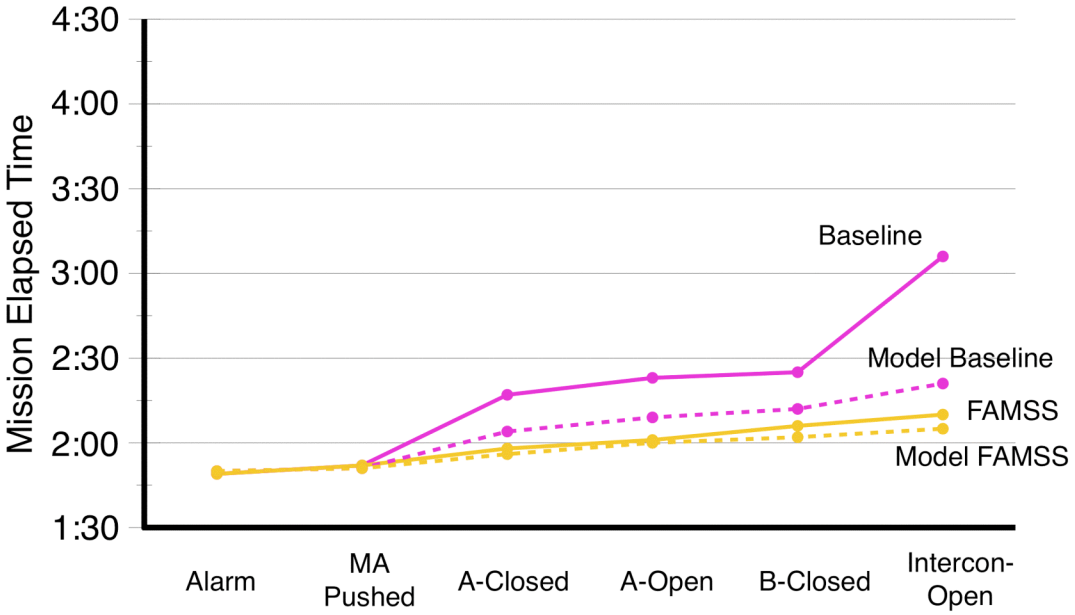


Figure 4.1. *Procedure completion times for the isolatable helium leak for the fastest participant and the APEX-GOMS model. Baseline Condition is in pink; FAMSS condition is in yellow. Dotted lines are model predictions; solid lines are the results from the participant.*

sec compared to 21 sec in FAMSS, a 55 second (72%) reduction. The predicted values from the APEX-GOMS model were 38 seconds to complete the malfunction in Baseline and approximately 19 seconds in FAMSS, an 18 second (50%) reduction. Given our assumptions, most astronauts would show FAMSS-related savings somewhere between these two values, or between 18 and 55 seconds, corresponding to somewhere between a 50% and 72% reduction.

With certain assumptions, then, human performance modeling can help us generalize our findings, and predicted FAMSS benefits, to populations of operators (like highly trained astronauts) outside the population (retired airline pilots) that served as the source of our sample. Working the other way, eye movements are an invaluable source of data to further develop the APEX-GOMS model to make it more realistic and better able to predict human performance in future cockpits. Because they provide direct data on the durations of individual actions and individual acts of information acquisition, fixation durations can be used to derive much more accurate estimates for the mean duration of many of the behavioral primitives included in the APEX-GOMS model, and critically, also provide the distribution (variance) around the means. In turn, this variance could be incorporated into the models to make the durations for the behavioral primitives in, for example, Appendix B, stochastic rather than fixed. By concatenating stochastically derived durations for the primitives, we could build and run Monte Carlo simulations of cockpit operations with realistic levels of variability.

4.5 Future Directions

4.5.1 Expanding FAMSS Capabilities

The shuttle operations paradigm has been refined over 25 years of flight. Each onboard task that the crew is required to accomplish onboard is developed, perfected, and practiced many times before flight. Simultaneously, Mission Control Center (MCC) ground controllers also learn, practice and perfect their tasks of systems monitoring and assisting with failure diagnoses. Though it would be desirable to reduce training time or introduce automation to lessen crew and ground controller workload, for the most part, the paradigm works and leads to successful missions.

Nevertheless, we previously noted that the circumstances of next-generation vehicle missions will require changes to the current operational paradigm. For shuttle missions, MCC is staffed with a multitude of controllers, supporting scientists and other support personnel. Although we envision that staffing levels will be comparable for VSE missions, communication delays will preclude some of the forms of real-time ground support that are currently available to the crew. A fault management support system could provide considerable assistance with more autonomous operations. However, FAMSS would need additional features and capabilities to handle:

1. No direct access to necessary parameters.
2. No automated switch throws.
3. Root cause uncertainty.
4. Multiple procedures applicable to the same root cause.
5. Conflicting actions in procedures under multiple-malfunctions condition.
6. Unforeseen malfunctions.
7. Variable Autonomy.

A comprehensive fault management support system needs to incorporate methods and displays to deal with each of these issues. We do not propose complete solutions here, but primarily enumerate the issues. Before describing these issues, we discuss some of the limitations of the study and future directions that may address them.

*No direct access to systems parameters***.** One of the core assumptions of the FAMSS concept is that onboard computers will have real-time access to all sensor data required to evaluate logical FDF expressions, fully automating the process of navigating through FDF procedures. What if that is not the case? Depending on the percentage of parameters available, the ramifications to FAMSS could be significant, but not insurmountable. FAMSS could require either minor modification, requiring the crewmember to input data that is not electronically accessible, or major modification, requiring a redesign to be more in line with the electronic checklists currently implemented on some glass-cockpit aircraft. If only a few parameters are not accessible, the basic features of FAMSS could remain unaffected. The redundant presentation of FDF instructions as both text and embedded into system schematics is still feasible. The automatic navigation to the next appropriate instruction may be affected and may require increased interaction with the crewmember to evaluate some logical expressions generated by FAMSS to assist it in filling in missing values. Note that navigation to the correct area of the procedure could still be automated once the crewmember provides a “true” or “false” answer to that logical expression.

No automated switch throws. Similar to access to sensor data, FAMSS assumes electronic access to mode control switches and the like. Automating switch throws has a number of advantages, including helping the crew locate the correct switch, enabling remote switch throws from ground controllers, and freeing up cockpit real estate. If switches cannot be electronically commanded, automatic mode reconfigurations would not be feasible. There are other possible solutions, even in this case, however. For example, the schematic section of the Fault Management Display could be modified to depict where in the cockpit the switch is located. Similarly, augmented reality techniques could highlight the location of the actual switch by illuminating it (some aircraft have already implemented this), using three-dimensional audio signals, or using tactile feedback.

*Root Cause Uncertainty***.** A third FAMSS prerequisite is root-cause fault determination. In its current form, FAMSS works only for cases in which the root cause is completely deterministic. The proposed concepts do not necessarily apply when the root cause is uncertain. One method of dealing with uncertainty is to eliminate it during vehicle design by adding redundant sensors throughout the systems. Two problems with this approach are the cost (in terms of additional weight) and the reliability of sensors (they tend to be much less reliable than the systems they are sensing). Even with redundant sensors, if enough sensors fail, the information received from the remaining instrumentation may be insufficient to disambiguate between the possible root causes of the symptoms. Under communication delay conditions, if an automated diagnosis system were unable to provide a single root cause, the crew would be responsible for disambiguating the problem. In a parallel effort, NASA Ames researchers are developing an ISHM Decision Analysis Tool (IDAT) that supports the crew in disambiguating multiple root causes and selecting the proper procedure to mitigate a problem (Spirkovska, 2006). IDAT contains concepts for the type of information necessary to adequately describe ambiguous symptoms to the crew, ways to provide the possible multiple root causes, what information about each possible root cause is important, and how to present malfunction impacts.

*Multiple procedures applicable to the same root cause***.** On a related note, different procedures may apply for the same problem under a different flight phase or conditions. For example, during the ascent phase, the procedure for handling a leak may be to interconnect to another system that uses the same propellant, whereas during the orbit phase, the procedure for the same leak may be to troubleshoot further and determine the cause of the leak. In the IDAT effort, we are exploring how to utilize decision analysis techniques to assist the crew in properly navigating to the correct (context-dependent) procedure.

Conflicting actions in procedures under multiple-malfunctions condition. A third challenge is how to handle situations in which a second malfunction occurs while the first is being worked or multiple malfunctions occur simultaneously. The current method is to work the highest-priority procedure first and finish the procedure completely before starting on the next highest-priority procedure. However, it is possible that the current procedure may change the state of the affected systems in a way that makes it impossible to recover from the second malfunction. It may be desirable to interleave the procedures in real-time to best recover from both (or all) malfunctions. Assuming a system can be developed to perform this interleaving, the FAMSS concept needs to be extended to support more complex forms of crew-system interaction. One issue is that there would not be a paper backup for interleaved procedures, so a process would

need to be developed to deal with interleaved procedures under FAMSS loss conditions (e.g., due to an electrical bus failure). Another issue is how to make the interleaving transparent to the crew – that is, what needs to be shown to the crew so that they can verify – even when they trust – that the automation is performing the interleaving correctly.

*Unforeseen malfunctions***.** The procedures in the FDF were developed in part as a result of a failure modes, effects and criticality analyses performed during design and in part to deal with unforeseen problems discovered during mission training and operations. In shuttle operations, when unforeseen problems are detected, the ground controllers devise a recovery procedure as quickly as possible. They then send that procedure up to the crew. After stepping through the procedure together (MCC and crew), the crew can execute the new procedure. MCC and the crew monitor the consequences of the procedure and intervene as appropriate by possibly changing the procedure to achieve the desired state. FAMSS needs to consider how to support the crew in dealing with these unforeseen situations when in a delayed-communication-with-ground condition. Some issues that must be considered include what type of information is necessary to reassure the crew that the problem can wait for ground controller involvement; if the problem has to wait, what can be done to safe the vehicle against possible escalations; if a procedure must be devised immediately, what information might help the crew in doing so or verifying that an automated system has devised a good option.

Variable autonomy. Under conditions similar to those of our study, the choice of the intermediate level four on the Sheridan-Verplank scale of human-machine function allocation appears to be well supported. However, in an actual mission, especially one far from Earth, level four may not always be the best choice. A lower or higher level of automation may be justified by such factors as the criticality of the malfunction, the crew's current activities and the crew's stress level. For example, if the crew is asleep when a minor malfunction occurs, it may be beneficial to have FAMSS automatically execute the recovery procedure and inform the crew of its activities once they wake. Likewise, if the crew is working a critical malfunction procedure, it may be beneficial to have FAMSS automatically execute the procedure for a malfunction that has a short window in which recovery is possible. If the window for recovery is long, the best option in this case may be to suppress the malfunction alarm until a more opportune time. Alternatively, if a malfunction procedure is highly critical, the crew may choose to see each step of the procedure, without any automatic navigation assistance from FAMSS. Or if the recovery procedure for a malfunction can be easily verified, the crew may choose to give "permission" at a higher-level than the instruction level proposed by FAMSS. In each case, depending on conditions, it may be useful to flexibly adjust the Crew-FAMSS functional allocation in order to decrease crew workload. The crew could make this determination themselves, of course, based on their own subjective determination of their current state, capabilities and needs. Alternatively, sophisticated monitoring tools are currently under development that assess crew activities and performance patterns in real time, and make automatic determinations of crew workload and current cognitive function (Schmorrow & Kruse, 2004; Raley, Stripling, Kruse, Schmorrow, & Patey, 2005). The threshold for what type of fault management decisions requires crew input could be automatically raised or lowered depending on the real-time assessments of crew state and current workload. Either way, to help the crew maintain high situation awareness, FAMSS will need to inform the crew of all actions it has taken on their behalf.

4.6 Concluding Remarks

Diagnosing and working malfunctions on a vehicle as complex and dynamic as a spacecraft is a demanding activity that can strain human performance capabilities to their limit. Automating the activities that contribute to these demands can result in dramatic improvements to the performance of the crew-vehicle system. These improvements are not as critical when the ground is available to assist with the more difficult aspects of the operations, such as making root-cause failure determinations. If we want to design a more autonomous vehicle, such automation is virtually mandated.

As evidenced by the content of this report, designing a fault management support system to enable a more autonomous concept of fault management operations is not a trivial task. It goes beyond what color a widget should be, or how to indicate that a procedure has been completed. There are many matters that need to be managed. In this evaluation, we find compelling evidence that an integrated approach to fault management support that incorporates automated root cause determination, automated procedure navigation, and automated switch-throws, and consolidates and supports human-automation interactions through a single display format, collectively provides substantial improvements to fault management performance with reduced workload and better situation awareness.

Designers of C&W and fault isolation and recovery operations on VSE spacecraft can use these results to help define operational targets for fault management effectiveness and efficiency. However, cockpit display real estate on these spacecraft is going to be at a premium, and designers are going to need much more specific guidelines as to what features of a proposed operational concept and supporting displays are beneficial, detrimental, or inconsequential, and whether users make use of design features the way designers intended. Our experience with the FAMSS evaluation suggests that these more specific requirements can be met only with human-in-the-loop evaluations that measure and analyze performance at different levels of behavioral specificity and temporal granularity, from subjective ratings of workload and situation awareness (which provide an assessment of performance across an entire run and are sensitive only to general aspects of cockpit design and operational difficulty), through measures such as response accuracy and latency, that work at intermediate levels of temporal and behavioral granularity, down to measures such as eye movements and human performance modeling tools that capture the most basic elements of human behavior at the smallest temporal scales.

5 References

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Boorman, D. (2000, Sep. 27-29). *Reducing flight crew errors and minimizing new error modes with electronic checklists.* In Proceedings of the International Conference on Human-Computer Interaction Aeronautics, Toulouse, France.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chuah, M. C., John, B. E., & Pane, J. (1994, Apr. 24-28). *Analyzing graphic and textual layouts with GOMS: Results of a preliminary analysis.* In Proceedings of the ACM CHI 94 Human Factors in Computing Systems Conference, Boston, MA.

Duchowski, A. T. (2003). *Eye Tracking Methodology: Theory and Practice*. London, UK: Springer-Verlag.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R., & Kiris, E. O. (1995). Out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied, 6*(4), 322-335.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA-TLX (task load index): results of empirical and theoretical research. In P. S. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier Science Publishers B. V.

Hayashi, M., Huemer, V., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005a, Sep. 26-30). *Effects of the space shuttle cockpit avionics upgrade on crewmember performance and situation awareness.* In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, FL.

Hayashi, M., Beutter, B., & McCann, R. S. (2005b, Oct. 10-12). *Hidden Markov Model analysis for space shuttle crewmembers' scanning behavior.* In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI.

Huemer, V., Hayashi, M., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005a, Sep. 26-30). *Characterizing scan patterns in a spacecraft cockpit simulator: expert vs. novice performance.* In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, FL.

Huemer, V. A., Matessa, M. P., & McCann, R. S. (2005b, Oct. 10-12). *Fault management during dynamic spacecraft flight: effects of cockpit display format and workload.* In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI.

John, B. E. (1990, Apr. 30-May 4). *Extensions of GOMS analyses to expert performance requiring perception of dynamic visual and auditory information.* In Proceedings of the ACM CHI 90 Human Factors in Computing Systems Conference, Seattle, WA.

John, B. E., & Gray, W. D. (1992, May 3-7). *GOMS analyses for parallel activities.* In Proceedings of the ACM CHI 92 (Tutorial materials), Monterey, CA.

Kasarskis, P., Stehwien, J., Hickox, J., & Aretz, A. (2001, May 5-8). *Comparison of expert and novice scan behaviors during VFR flight.* In Proceedings of the 11th International Symposium on Aviation Psychology, Columbus, OH.

Malin, J. T., Schrechenghost, D. L., Woods, D. D., Potter, S. S., Johannesen, L., Holloway, M., et al. (1991). *Making intelligent systems team players: case studies and design issues, Vol. 1: Human-computer interaction design* (NASA TM No. 104738). Houston, TX: NASA.

Malin, J., Kowing, J., Schreckenghost, D., Bonasso, P., Nieten, J., Graham, J. S., Fleming, L. D., MacMahon, M., & Thronesbery, C. (2000). Multi--agent diagnosis and control of an air revitalization system for life support in space. *2000 IEEE Aerospace Conference CP*

Matessa, M., & Remington, R. (2005a, Sep. 26-30). *Eye movements in human performance modeling of space shuttle operations.* In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, FL.

Matessa, M., & Remington, R. (2005b, Oct. 0-12). *Reusable templates of human performance in space shuttle procedures.* In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI.

McCandless, J. W., McCann, R. S., Berumen, K. W., Gauvain, S. S., Palmer, V. J., Stahl, W. D., et al. (2005, Sep. 26-30). *Evaluation of the space shuttle cockpit avionics upgrade (CAU) displays.* In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, FL.

McCandless, J. W., McCann, R. S., & Hilty, B. R. (2003, Oct. 10-17). *Upgrades to the caution and warning system of the space shuttle.* In Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, Denver, CO.

McCann, R. S., & McCandless, J. W. (2003, Dec. 1-5). *Human-machine teaming for dynamic fault management in next-generation launch vehicles.* In Proceedings of the Joint Army-Navy-NASA-Air Force (JANNAF) 3rd Modeling and Simulation Subcommittee Meeting, Colorado Springs, CO.

McCann, R. S., & Spirkovska, L. (2005, Nov. 7-10). *Human factors of integrated systems health management on next-generation spacecraft.* In Proceedings of the First International Forum on Integrated System Health Engineering and Management in Aerospace, Napa, CA.

Moody, J., & Darken, C. J. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation, 1*(2), 281-294.

National Transportation Safety Board (NTSB). (1989). *Aircraft Accident Report. Northwest Airlines, Inc., McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus Michigan, August 16, 1987* (No. AAR-88/05, Adopted 04/28/1989, DCA87MA046, File No 758).

Raley, C., Stripling, R., Schmorrow, D., Patrey, J., & Kruse, A. (2004, Sep. 20-24). *Augmented cognition overview: Improving information intake under stress.* In Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society, New Orleans, LA.

Roscoe, A. H. (1984). Assessing pilot workload in flight. In *Advisory Group for Aerospace Research & Development Conference Proceedings No. 373: Flight Test Techniques*. Neuilly-sur-Seine, France: NATO.

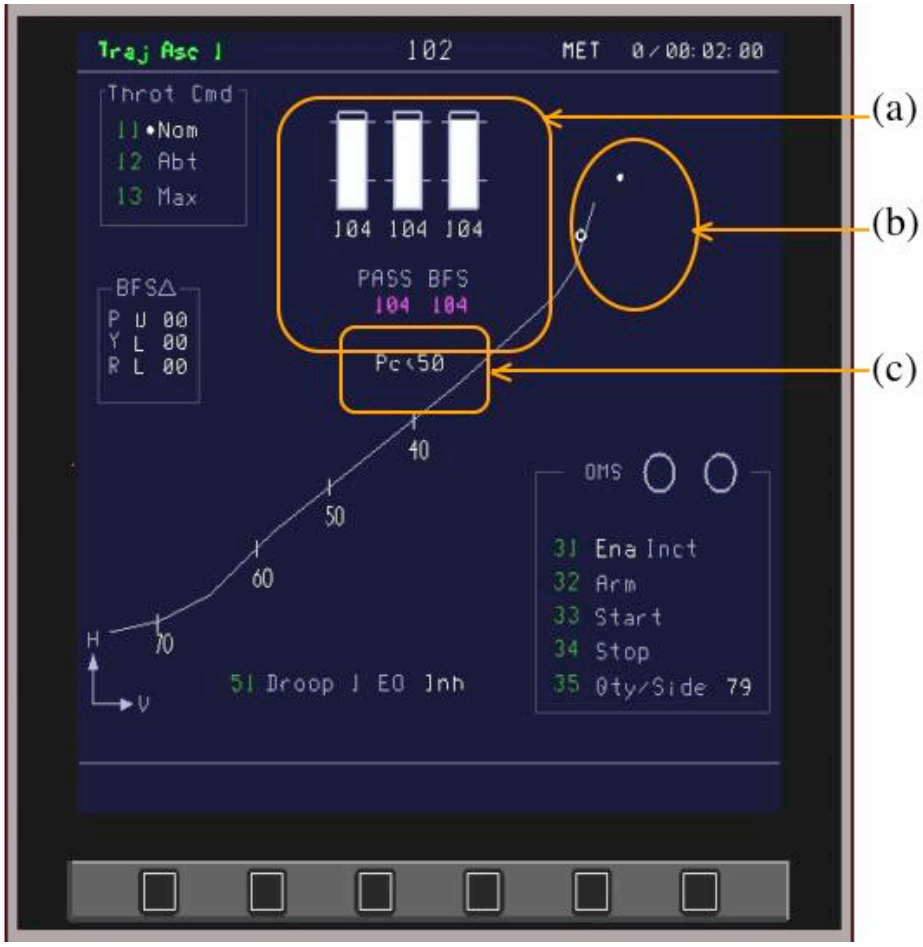
Scandura, P. A., & Garcia-Galan, C. A. (2004, Oct. 24-28). *A unified system to provide crew alerting, electronic checklists and maintenance using IVHM.* In Proceedings of the IEEE Digital Avionics Systems Conference, Salt Lake City, UT.

Schmorrow, D., & Kruse, A. (2004). Augmented Cognition. In W. S. Bainbridge (Ed.), *Berkshire Encyclopedia of Human-Computer Interaction*: Berkshire Publishing Group.

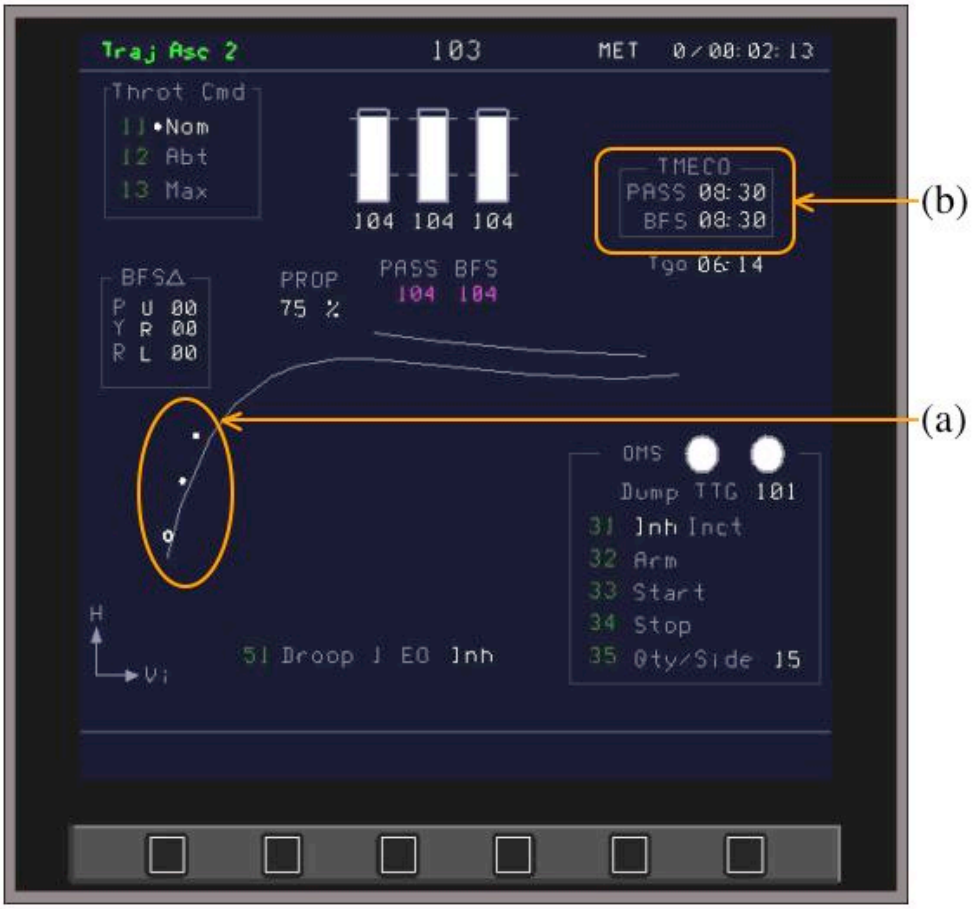
Spirkovska, L. (2006). *ISHM decision analysis tool: Operations concept* (NASA TM No. 2006-213481). Moffett Field, CA: NASA.

Appendices

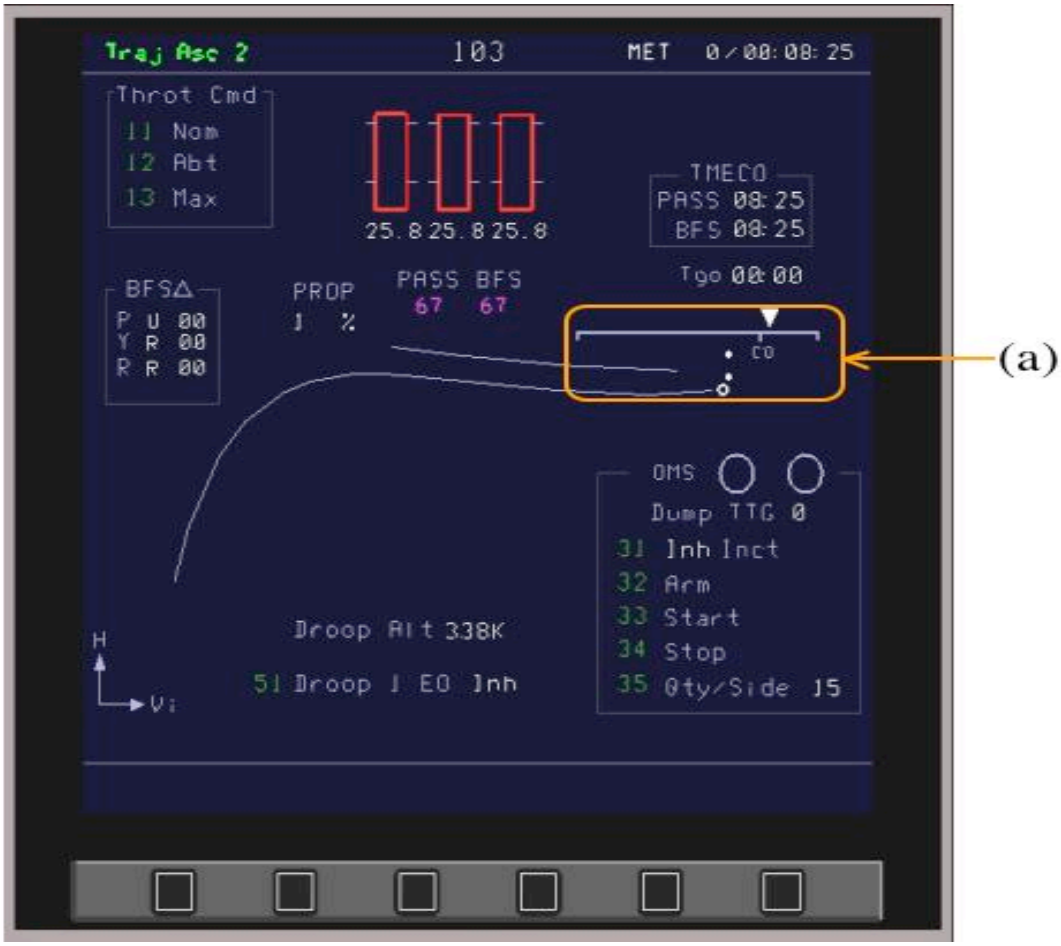
Appendix A: Displays used during nominal monitoring utilizing the PAHUEE scan.



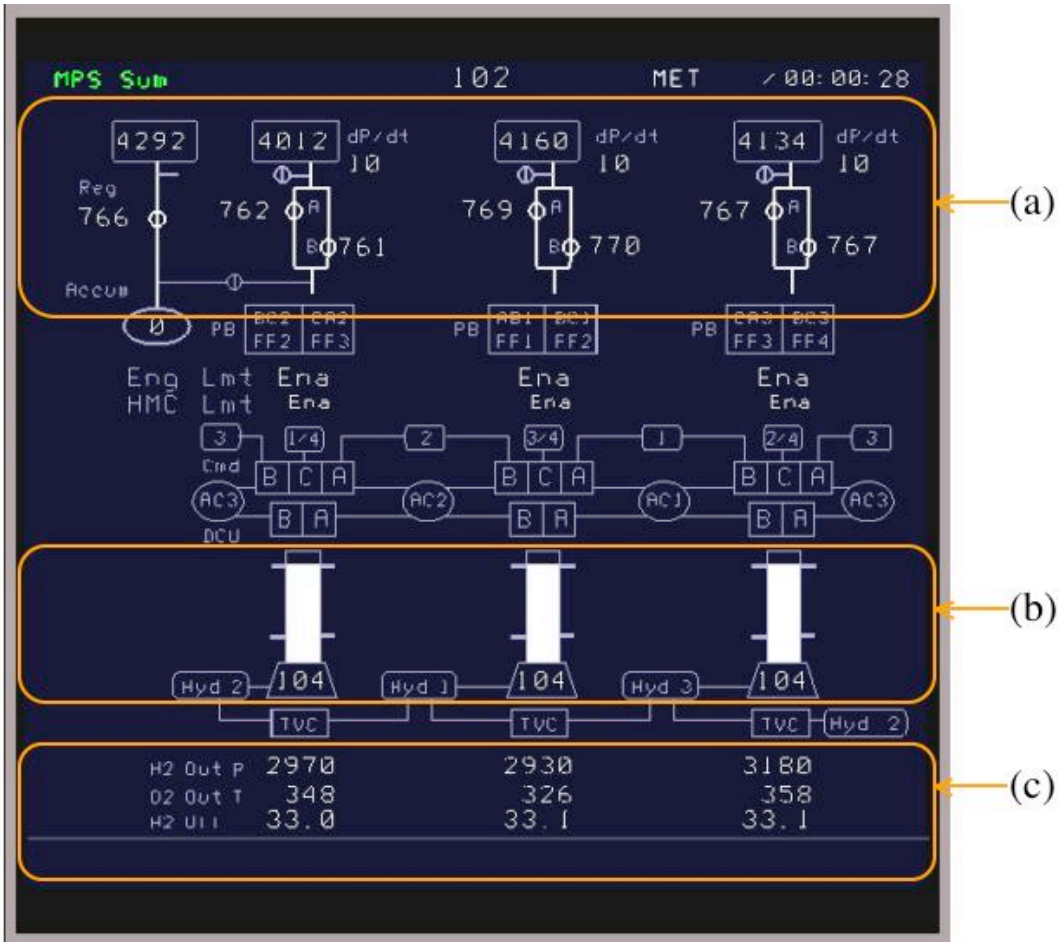
Display A.1: CAU Ascent Trajectory (Traj Asc 1), just before SRB separation. Arrow (a) indicates the main engine thrust indicators. Arrow (b) indicates the Solid Rocket Booster chamber pressure. Arrow (c) indicates the pitch (current and predicted in 20 sec.). Provides pitch information for PAHUEE scan.



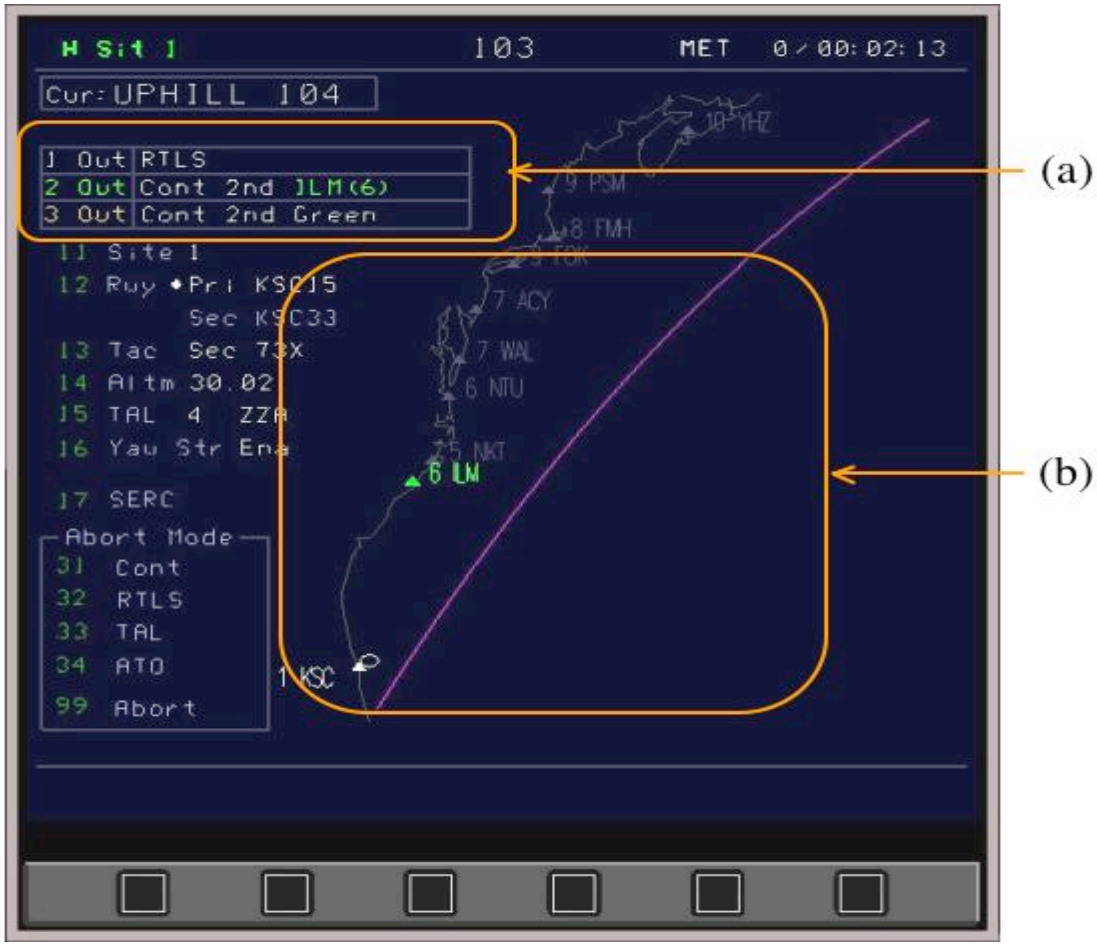
Display A.2: CAU Ascent Trajectory Display after SRB separation (Traj Asc 2). Arrow (a) indicates pitch (current and predicted in 30 sec, and in 60 sec). Arrow (b) indicates PASS- and BFS-computed time to main engine cutoff (MECO) values. Display provides pitch information for PAHUEE scan.



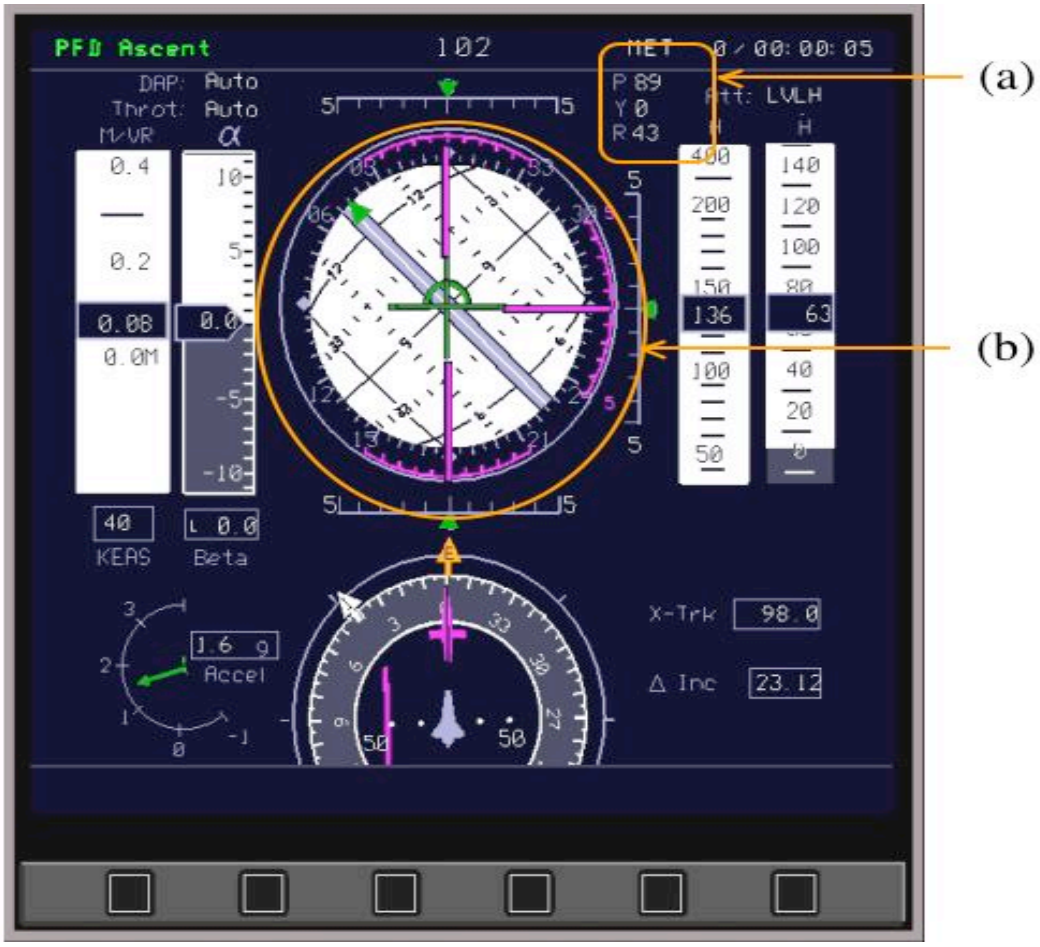
Display A.3: CAU Ascent Trajectory Display (Traj Asc 2) prior to MECO. Arrow (a) indicates MECO velocity cutoff bug. Display provides pitch information for the PAHUEE scan.



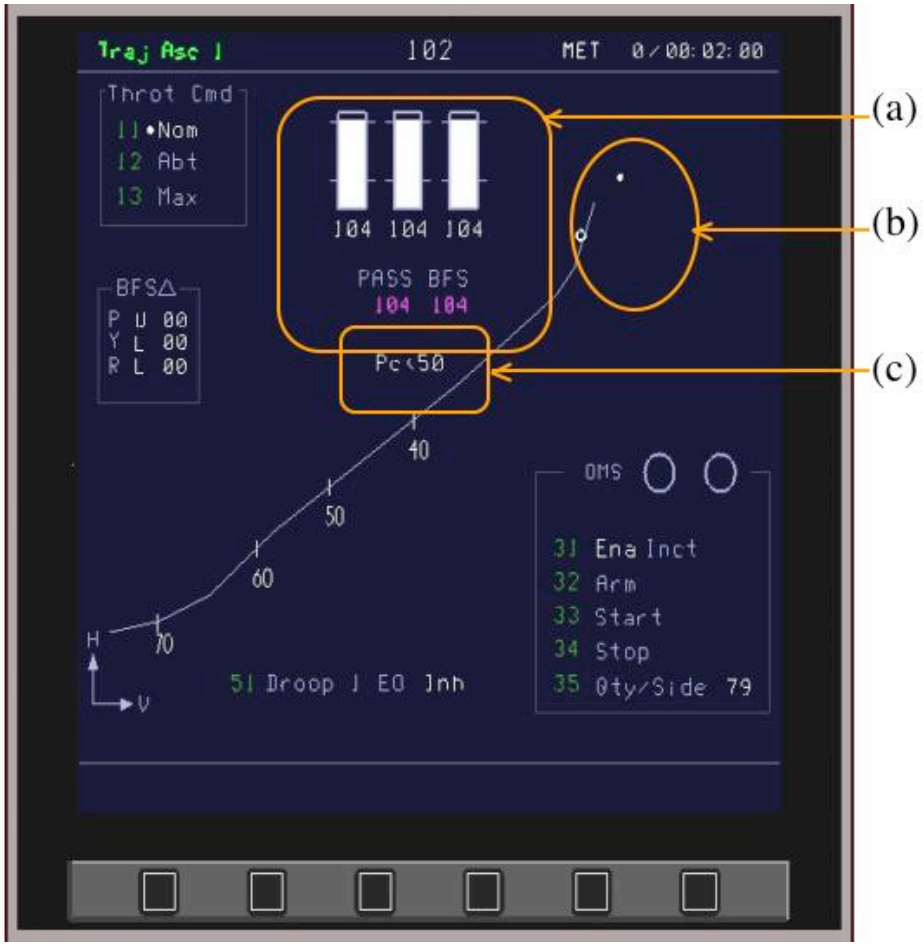
Display A.4: CAU Main Propulsion System Summary Display (MPS SUM) during nominal run (see Figure 1.3 for example MPS SUM in off-nominal run). Arrow (a) indicates helium supply system information, arrow (b) indicates MPS thrust information, and arrow (c) indicates external tank ullage pressures. Provides main engine helium supply system information and ullage information for the PAHUEE scan.



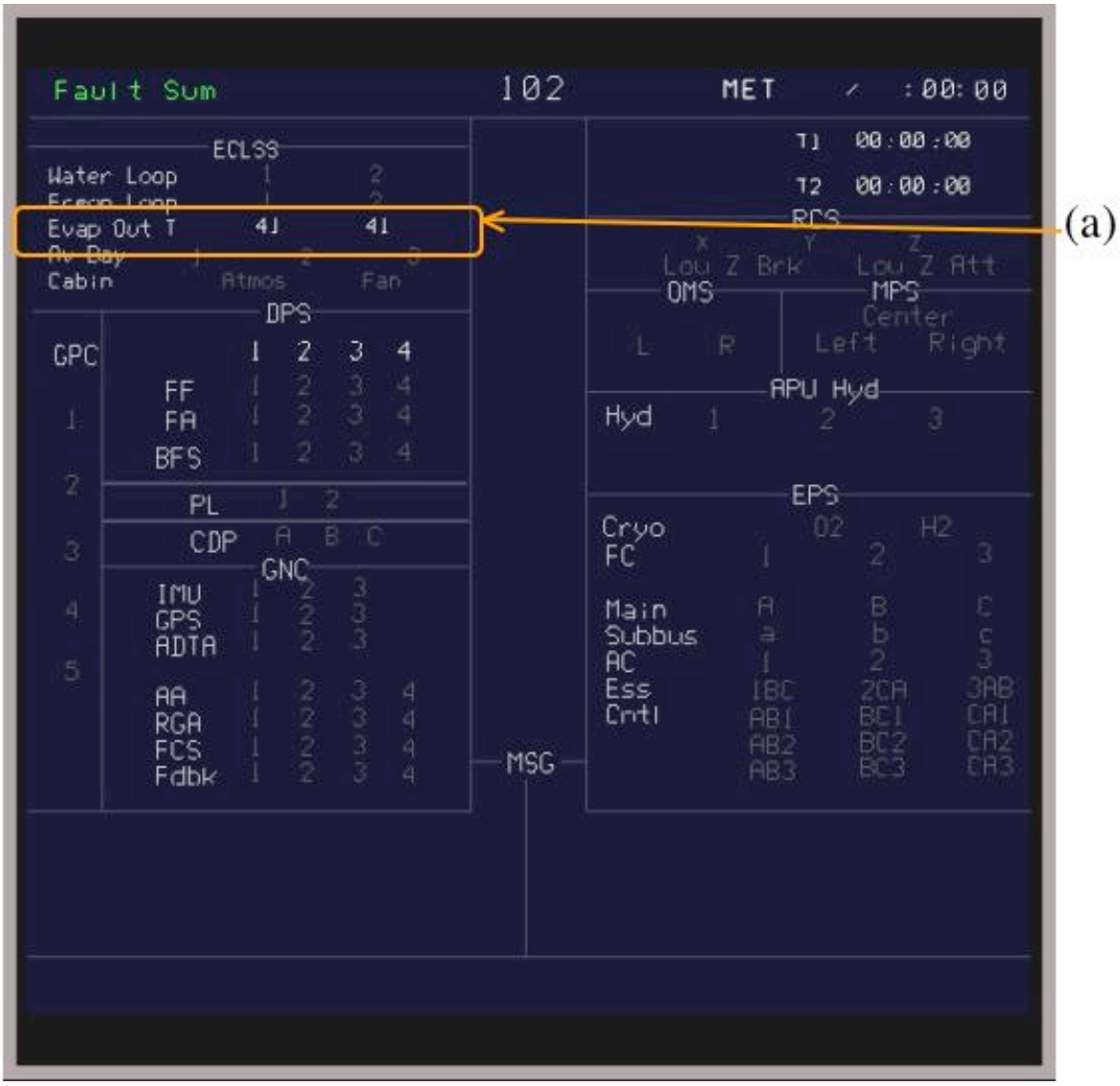
Display A.5: CAU Horizontal Situation (H Sit). Arrows indicate sections containing information concerning current abort options. Provides abort information for PAHUEE scan.



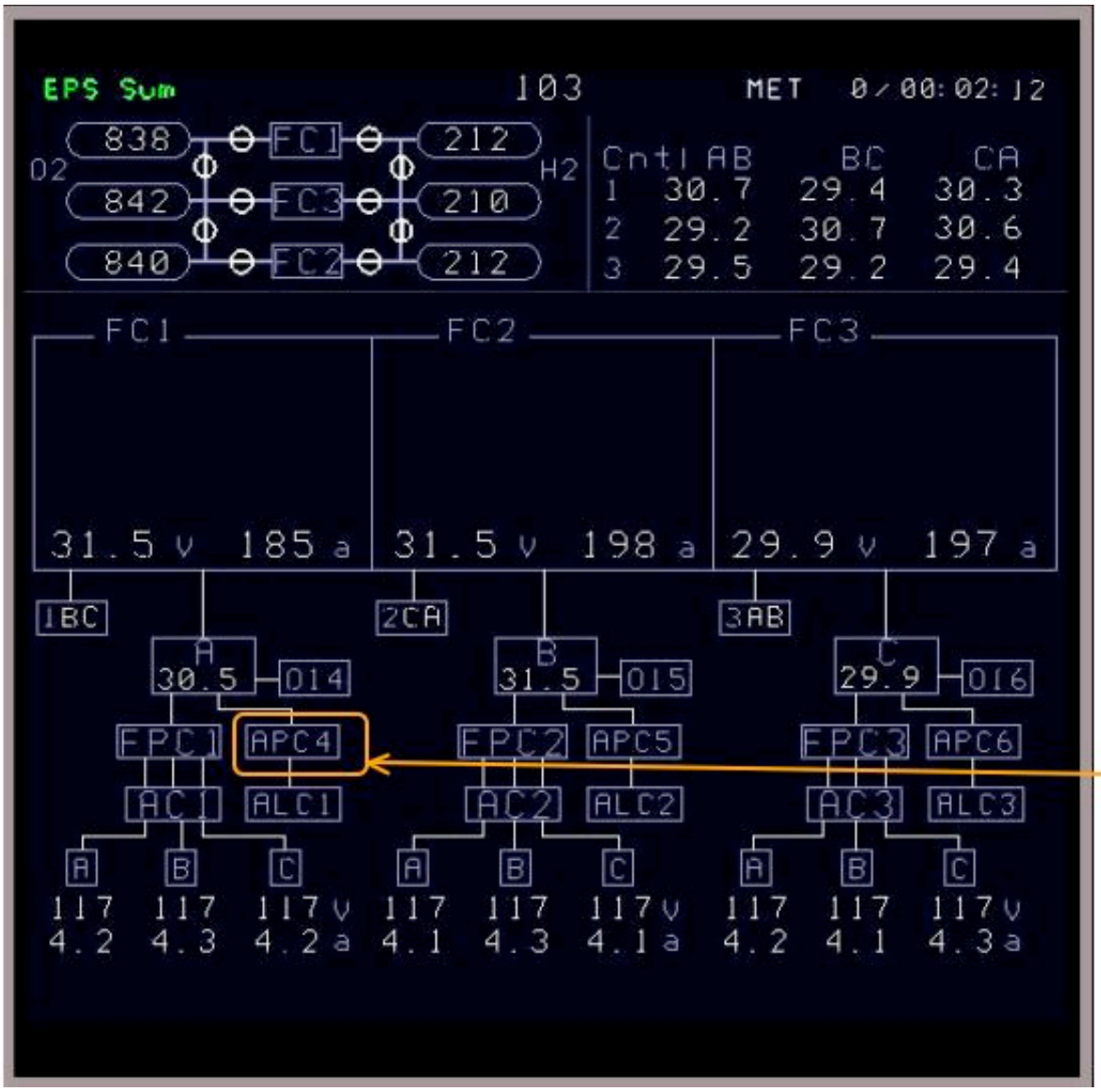
Display A.6: CAU Composite Attitude Director Indicator/Horizontal Situation Indicator (ADI/HSI). Arrow (a) indicates numeric attitude readings (pitch, roll, and yaw). Arrow (b) indicates the attitude ball, used to display pitch and roll. Provides pitch and abort information for PAHUEE scan. In Baseline condition, visual Master Alarm light is located above the upper right corner of the display, above MET.



Display A.7: CAU Ascent Trajectory (Traj Asc 1), just before SRB separation. Arrow (a) indicates the main engine thrust indicators. Arrow (b) indicates the Solid Rocket Booster chamber pressure. Arrow (c) indicates the pitch (current and predicted in 20 sec.). Provides pitch information for PAHUEE scan.



Display A.8: CAU Fault Summary (Fault Sum). Arrow (a) indicates Evap Out T values used during PAHUEE scan.



Display A.9: CAU Electrical Power System Summary (EPS SUM). Provides EPS system status during the PAHUEE scan. Arrow (a) shows location of APC4 subbus in distribution assembly section of display.

Appendix B: Behavioral Primitives and Temporal Predictions for Selected Malfunctions

APC4 Subbus Failure

Baseline Condition

hear alarm	50		
orient Master Alarm	470		
see & press alarm	720	alarm	1240
orient Fault Summ	470		
read 'APU 3 spd low	1020		
read 'MPS LH2/LO2 Ull P	1410		
see subbus a (red)	470		
orient DPS/EPS	470		
press tab DPS	720	DPS tab	4560
locate & confirm APC 4 failure	470		
see/confirm APU HYD RPC B (red)	470		
see/confirm APU 3 speed low 0 (yellow)	470		
orient to MPS	470		
see/confirm Ullage C low (Yellow arrow)	470		
see/confirm MPS He C Isol A (red)	470		
orient to Ack panel	470		
press Ack to clear msg.	720		
grab FDF	720		
find Mal page	2350		
read 'MN BUS UNDER V/FC V	1410		
read 'AC Volt ...	470		
read 'BUS TIE...	470		
read 'Subbus [APC4(5,6) or ALC1(2,3)]	2350		
read 'Do not isolate MPS He C(L,R)	2350	Malf. Handled Total time	14130 **19930**

FAMMS Condition

hear alarm	50		
orient AFMS	470		
see & press alarm	720	Alarm	1240
Read Mal Msg. 'APC 4 failure	1410		
Conf. Mal on schematic	470		
orient DPS/EPS	470		
press EPS	720	EPS tab	3070
see APU HYD RPC B (red)	470		
see APU 3 speed 0 (yellow)	470		
orient MPS	470		
see Ullage C low (yellow)	470		
see MPS He C Isol A (red)	470		
Orient AFMS	470		
Read 'Do not Isolate MPS He C(L,R)	2350	Handled Total time	5170 **9480**

GPC Fail-to-Synch

Baseline Condition

hear alarm	50		
orient Master Alarm	470		
see & press alarm	720	Alarm	1240
orient to Fault Sum	470		
read Mal. Msg 'GPC 4 fail to 123	1410		
orient uperhead panel	470		
see/conf. GPC 4 fail	470		
grab FDF	720		
find mal page in FDF	2115		
read 'Pass GPC fail	940		
read 'Aff FCS CH - OFF	1410		
orient FCS CH 4 panel	470		
Switch FCS 4 OFF	720	FCS-Off	9195
orient to DPS/EPS	470		
press tab DPS	720		
locate & confirm FCS CH 4 on DPS	470		
orient to FDF	470		
read 'If two GPC/FA/FCS CHI	1880		
read 'If FTS and accessible - STBY,HLT	2350		
Orient to panel O6	470		
Switch Mode STBY	720	Mode-STBY	7550
orient DPS	470		
locate & confirm Mode - STBY	470		
Orient to panel O6	470		
Switch Mode HLT	720	Mode-HLT	2130
orient DPS	470		
locate & confirm Mode - HLT	470		
		Total time	21055

FAMMS Condition

hear alarm	50		
orient AFMS	470		
see & press alarm	720	Alarm	1240
Read Mal. Msg. 'GPC 4 fail to Sync.	1410		
orient uperhead panel	470		
see/conf. GPC 4 fail	470		
orient AFMS	470		
press GPC tab	720		
Read 'FCS Channel 4 - Off Accept	1410		
press Accept	250	FCS 4 OFF	5200
orient DPS	470		
Confirm FSC 4 - OFF	470		
orient AFMS	470		
Read 'GPC mode 4 stby Accept	1880		
press Accept	250	Mode STBY	3540
orient DPS	470		
Confirm Mode 4 STBY	470		
orient AFMS	470		
Read 'GPC mode 4 HLT accept	1650		
press Accept	250	Mode HLT	3310
orient DPS	470		
Confirm Mode 4 HLT	470		
		Total time	14230

APC4 Subbus Failure

Baseline Condition				FAMMS Condition			
hear alarm	50			hear alarm	50		
orient Master Alarm	470			orient AFMS	470		
see & press alarm	720	alarm	1240	see & press alarm	720	Alarm	1240
orient Fault Summ	470			Read Mal Msg. 'APC 4 failure	1410		
read 'APU 3 spd low	1020			Conf. Mal on schematic	470		
read 'MPS LH2/LO2 Ull P	1410			orient DPS/EPS	470		
see subbus a (red)	470			press EPS	720	EPS tab	3070
orient DPS/EPS	470			see APU HYD RPC B (red)	470		
press tab DPS	720	DPS tab	4560	see APU 3 speed 0 (yellow)	470		
locate & confirm APC 4 failure	470			orient MPS	470		
see/confirm APU HYD RPC B (red)	470			see Ullage C low (yellow)	470		
see/confirm APU 3 speed low 0 (yellow)	470			see MPS He C Isol A (red)	470		
orient to MPS	470			Orient AFMS	470		
see/confirm Ullage C low (Yellow arrow)	470			Read 'Do not Isolate MPS He C(L,R)	2350	Handled	5170
see/confirm MPS He C Isol A (red)	470					Total time	**9480**
orient to Ack panel	470						
press Ack to clear msg.	720						
grab FDF	720						
find Mal page	2350						
read 'MN BUS UNDER V/FC V	1410						
read 'AC Volt ...	470						
read 'BUS TIE...	470						
read 'Subbus [APC4(5,6) or ALC1(2,3)]	2350						
read 'Do not isolate MPS He C(L,R)	2350	Malf. Handled	14130				
		Total time	**19930**				

Isolatable Helium Leak

Baseline Condition				FAMMS Condition			
hear alarm	50			hear alarm	50		
orient to Master Alarm	470			orient AFMS	470		
see & press alarm	720	alarm	1240	see & press alarm	720	Alarm	1240
orient Fault Summ	470			Read Mal Msg. 'BFS R MPS He P Low	1410		
read 'MPS He P"	940			orient MPS	470		
orient MPS	470			Confirm Malf.	470		
confirm MPS Malfunc.	470			orient AFMS	470		
grab FDF	720			Read 'R He Isol A - Close Accept	2350		
find Mal page	2115			press Accept	250	Isol A - close	5420
read '/dp/dt	470			see dp/dt on Schematic	470		
orient MPS	470			Read 'R He Isolation A - open Accept	2350		
check dp/dt	470			orient Asc Traj	470		
orient to FDF	470			SRB check Pc meter	470		
read 'if after MECO - 60	940			SRB identical TMECO	470		
orient Asc_traj	470			orient AFMS	470		
look at MET	470			press Accept	720	Isol A - open	5420
SRB check Pc meter	470			Read 'R He Isolation B - Close Accept	2350		
SRB check identical TMECO	470			press Accept	250	Isol B- close	2600
orient to FDF	470			check dp/dt	470		
read 'if He reg P |^ or |v	1410			Read 'R He Interconnect - IN open Accept	2820		
orient to MPS	470			press Accept	250	Inter open	3540
look at Reg P	470			confirm Malf. Solved in schematic	470		
orient to FDF	470					Total time	**18690**
read 'otherwise	470						
read 'Aff He Isol A - CL	1410						
orient to panel	470						
see & press ISOL-A - CL	720	ISOL A -CL	16245				
orient to FDF	470						
read 'if no decr in dp/dt	1410						
orient MPS	470						
check dp/dt	470						
orient to FDF	470						

read 'aff He ISOL A - Op	1410		
Orient Panel	470		
see & press ISOL - A OP	720	ISOL A -OP	5890
orient FDF	470		
read 'Aff He ISOL B- CL	1410		
Orient Panel	470		
see & press ISOL - B CL	720	ISOL B - CL	3070
orient FDF	470		
read 'if no decr in dp/dt	1410		
orient MPS	470		
check dp/dt	470		
orient FDF	470		
read 'if any ENG failed	1410		
orient Asc_traj	470		
check Engines (red)	470		
orient FDF	470		
read 'if nonisolatable	1410		
read 'if isolated	940		
read 'Aff He I'CNT - IN OP	1410		
orient to Panel	470		
see & press Interconnect OP	720	Inter - OP	11060
		Total time	**37505**